AN URBAN MASS TRANSIT MODEL WHICH CONSIDERS DEMAND ELASTICITY, ROUTE STRUCTURE, AND PERCEIVED PASSENGER TRAVEL TIME

A THESIS

Presented to

The Faculty of the Division of Graduate Studies

by

W. Jackson Ritchie

In Partial Fulfillment

of the Requirement for the Degree

Master of Science in Operations Research

Georgia Institute of Technology

March, 1976

AN URBAN MASS TRANSIT MODEL WHICH CONSIDERS DEMAND ELASTICITY,

ROUTE STRUCTURE, AND PERCEIVED PASSENGER TRAVEL TIME

APPROVED:

G.P. Sharp, Chairman

R.L. Rardin

J.J. Jarvis

Date approved by Chairman: 31 Mar 76

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my advisor, Dr. Gunter Sharp. Not only did he provide the motivation for this research topic, but his keen insight proved invaluble in helping to solve the many conceptual problems I faced. His willingness to read my rough drafts and make constructive suggestions especially with such short notice at times, will long be remembered.

I would also like to thank the members of my Reading Committee Dr. J.J. Jarvis and Dr. Ron Rardin, who took the time to listen to me and always ask "why?".

The support given me by Dr. Robert Lehrer and Dr. W.W. Hines allowed me to pursue the Mater's degree and gain a great deal of satisfaction from working with them.

Finally I would like to thank my wife, Dana. Not only did she spend many nights into the wee hours of the morning typing and sleeping in my office, but most important, put up with me and nursed me throughout my bouts with "terminal fever". This thesis is at least one-half hers.

This thesis is dedicated to my parents, whom I always knew were behind us.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF ILLUSTRATIONS

LIST OF TABLES

SUMMARY

There has been a significant emphasis in the development of mass transit systems during the last decade. Most models being used currently for planning mass transit systems are too simplified to give useful results to designers and planners.

This research includes the construction of a descriptive model which will allow the planner to obtain passenger assignments considering the factors perceived travel time, demand elasticity and route structure. Also a solution procedure for the model is developed and tested.

The model is network based with modifications. The perceived travel time is a weakly increasing function of riders on the vehicle, the demand for an origin destination pair is a weakly decreasing function of perceived travel time and finally, a passenger on a vehicle cannot be forced off due to congestion.

The network is solved with an heuristic algorithm which finds shortest paths, calculates effective demand and assigns passengers in an iterative fashion.

The solution procedure is tested with several small, randomly generated test problems. The results of the tests indicate the feasibility of constructing and using such a model in the transit planning process.

CHAPTER I

INTRODUCTION

Background

There has been a significant emphasis in the development of urban mass transit systems during the last decade. This has been due to the increase in urban population and the congestion which results during peak travel times, and the recent energy crisis. With the automobile now being questioned as an appropriate vehicle for urban travel, it is expected that there will be a shift toward the use of mass transit.

As the new transit systems are designed to handle the greater demand, their complexity increases also. Most models being used currently for planning mass transit systems are too simplified to give useful results to designers and planners. Typically, these models relax the restrictions that certain paths may not have the capacity to serve their demand, that particular path lengths may become costlier to the traveler as the paths become crowded, or finally, that passengers observe and react to various characteristics of their possible paths (3,14,19).

Therefore, there is a need to model accurately the behavior of the passengers and their effects on the system. Research has recently been undertaken to describe the behavior of urban travelers; however, this information has not been incorporated into a working model of a system.

The transit system may be considered basically as a network in which several characteristics must be considered. First, each arc has two capacities; one is seating capacity and the other is standing room. Some passengers will want to stay on the system only if a seat is available, while other travelers will ride as long as there is someplace to sit or stand. These separate capacities are related also to the concept of "perceived time" by the passenger. Since a discomfort factor must be considered if the traveler must stand, his perceived time on the system is greater than the actual time. Usually this factor is assumed to be about two times greater than the actual time (4). As an example, suppose a passenger takes a trip that lasts ten minutes. If he could ride in a seat for the trip's duration his perceived time would be the same 10 minutes; however, if he were forced to stand his perceived time would then be twice the actual time, or 20 minutes. Another example of the use of perceived time is the choice a traveler makes between alternate paths, or between a path involving a route transfer and a through route.

The reasoning behind developing this concept of perceived time is that passengers possess a certain upper bound on the time they are willing to spend on the system from their origin to destination. If their time is greater than the upper bound, then they leave the system to travel, say, by automobile. Thus the passenger assignment problem cannot be treated as a typical multi-commodity, multi-path network problem in which the commodities always take their assigned paths, regardless of how long they are.

Another characteristic of the system is that there exists a route structure. Generally, a substantial number of passengers flow toward the "center" of the urban area and the routes usually reflect this flow. This structure then permits the origination of routes at the outer edges of the urban area and loading them starting at the outer edges. The main effect of this loading will be that the passenger who boards near the outer edge will get the seat while the traveler nearer the center of town may be forced to stand. Furthermore, the seated passenger usually retains his seat until his destination, unless he transfers onto a busy route. The network, then, cannot be formulated as a simple traffic assignment problem.

Another consideration of the behavior of the passengers is the elasticity of their demand. They do not interact optimally in this system. Certain passengers will ride the system if at all possible. These are the ones with a large upper bound on their perceived time due to lack of other transportation. Other potential riders, just the opposite, will never use the system if there is a remote possibility they will not get a seat. In between these two extremes are the passengers who will choose to leave the system depending on what they evaluate as their perceived time, including the consideration of sitting or standing.

Obviously, if all the above special characteristics of urban transit are to be considered in a model, a simple traffic assignment or multi-commodity network flow representation will not suffice. The solution procedure for the problem then will not be able to take advantage of special techniques developed for these common models.

Even if certain assumptions were relaxed in order to formulate the problem closer to a standard model, the magnitude of the network would render infeasible any exact solution procedure because of the amount of computation time required.

Purpose

Given the above situation concerning transit passenger models, the purpose of this research is as follows:

1. To construct a descriptive model which will allow the planner to obtain passenger assignments considering the factors described above--perceived travel time, demand elasticity, and route structure.

2. To develop an efficient solution procedure for the model.

3. To obtain computational experience with the model solution procedure for small test problems.

Method of Approach

The descriptive model to be developed will be a network-based assignment model with modifications. First, the path travel time will be a weakly increasing function of expected waiting time, passenger loads on vehicles, and vehicle transfers included in the path. Second, the demand for an origin-destination pair will be a weakly decreasing function of the ratio between path travel time and automobile travel time. Third, the assignment rules include the specification that once a traveler boards a vehicle, he stays on the vehicle until his desired transfer or destination point. That is, a passenger on a vehicle cannot be forced to get off due to a congestion effect. Moreover, if a traveler occupies a seat immediately after boarding a vehicle, he continues to occupy that seat while riding the vehicle.

This third item dealing with assignment rules, and the subsequent effect upon a passenger's travel time, leads to the one-pass loading procedure to obtain an initial solution. Passengers at the outer edges of the network, at the route ends, are loaded first based on perceived path travel times. The procedure then works inward toward the central area of the network. Following the initial solution, the travel demands are re-examined, and adjustments are made in assignments until some convergence criterion is satisfied.

Computational experience will be obtained by performing an experimental design, using randomly generated test problems.

It is expected that the results gained from this research will yield a more useful description of urban passenger behavior for transit planners, provide insight into possible solution techniques, and indicate areas which may benefit from further investigation.

CHAPTER II

LITERATURE SURVEY

Traditionally, the urban transit planning process has consisted of several sequential related phases from the generation of trips to the assignment of routes. This chapter will review this process, including the difficulties encountered in applications to urban systems, and note alternatives similar to this research effort which attempt to overcome these difficulties.

The Travel Demand-Forecasting Process

Since the 1950's the demand-forecasting process has consisted of five primary steps: trip generation analysis, captive modal split analysis, trip-distribution analysis, choice modal split analysis, and traffic assignment analysis. These steps are carried out sequentially until the forecasted demand is loaded on the urban network (12).

The purpose of the trip generation phase is to estimate the trip ends of particular type trips based on the analysis of traffic zones. These trip ends are divided into two distinct parts; the first being trip production, which applies to trips generated by residential areas, and the second being trip attractions, which deals with non-home activities such as work and shopping. These trip generations can be developed to reflect different characteristics of the urban systems being studied, such as social class or time of day.

Multiple regression techniques are most often used in this phase, particularly stepwise regression analysis. This allows the planner to try various combinations of factors, as numbers of workers in a household, until he is satisfied with the results obtained. These generated trips are then input to the captive modal split phase.

The modal split analysis is concerned with determining how the trips generated are distributed among captive users who depend solely on public transportation and choice users who can travel by private auto or transit. Earlier techniques for calculating the modal split made no distinction between captive and choice users, and did not consider behavioral aspects of the demand. New approaches have been developed which do account for the behavior characteristics, and separately consider captive and choice users, the latter by defining travel-cost relationships as choice criteria between nodes. It is these approaches which are also being used to determine modal splits for choice users after the distribution phase.

The trip distribution phase is concerned with combining the trip ends generated to form O-D pairs for both the captive and choice riders. The majority of transit planners have used a form of the gravity model to synthesize the trip-distribution matrices. The basic idea behind the gravity model is to weight the attractiveness of a zone compared to other zones. A critical factor in the weighting procedure is the determination of travel time functions between zones.

Linear programming techniques have recently been applied to the trip distribution problem (22). Here the objective minimized is the total travel time between O-D pairs. The results of this application have been good, but apparently they would be better if the demands were split into socio-economic classes to help justify the objective function.

The final phase of the travel demand forecasting procedure deals with the distribution of the demand over the network. Almost all assignment techniques are based on one of Wardrop's two principles (24):

1) All the paths used between any given origin and destination have equal travel time, and there is no unused path between the same origin and destination with a lesser travel time.

2) The average journey time is a minimum.

The three basic techniques used in this phase are all-or-nothing assignments, capacity restraint algorithms, and multipath assignments.

The all-or-nothing assignment techniques basically find the shortest path for each origin--destination pair and assign the associated demand to the respective paths. The underlying assumption in this approach is that each travel demand in the network will be served by an adequate capacity, and thus the arc travel times do not vary with the amount of flow.

The capacity restraint algorithm, however, does consider the effects of congestion on travel times. Generally, demand is assigned in either a step-wise or an iterative fashion to shortest paths, with the arc travel times updated appropriately. The iterations continue until some convergence is established or some predetermined cutoff criterion is reached. As in the distribution phase, linear programming has been applied to the assignment problem, using Wardrop's second principle as the basis for the objective function. Here arc capacities are enforced and total travel time is minimized (20).

There are several multi-path assignment procedures which have been used for the assignment phase. Graph theory and Markov processes are two of the most popular techniques, both of which utilize flow dependent arc travel times for limiting capacity (3).

Combining Distribution - Assignment

The major drawback with the five-phase process is that it is sequential; the process does not allow the demand to change based on the travel times (costs) determined in the assignment phase. Techniques have been developed to recalculate the demands by recycling back into the distribution phase after the assignments. The most successful approach, however, appears to be a linear programming formulation of the combined distribution - assignment phases.

Tomlin (21) has proposed such a solution procedure based on the Dantzig-Wolfe decomposition algorithm. The procedure is based on the linear programming approach mentioned earlier in the distribution and assignment phases. The subproblems consist of finding shortest paths with constant arc travel times and also generating demand. The master problem, using a linear flow-travel time relationships, is then optimized to give a new distribution and assignment.

Further investigation is being made into the mathematical programming formulation as the linear cost functions are being relaxed and non-linear programs similar to the linear programming approach are being developed and tried. The most promising of these new models has been formulated by Leblanc (13).

Florian et al. have devised a new approach to the problem by considering elastic demands, similar to the demand functions used in this research effort and non-linear travel time functions (7). They formulate the problem in a mathematical programming context and use a decomposition procedure which utilizes an available fixed demand algorithm to optimize the subproblems.

One major drawback of these procedures is the large amount of storage space required for large networks. This aspect reduces their application to small problems at the present.

Networks with Routes

Up to this point all the literature reviewed has dealt with assigning demand to traffic networks. Until recently, the assignment phase in networks with route structure, such as public transit systems had been ignored, as the assumption was made that not enough alternative paths existed to make the effort worthwhile. As more large urban areas are increasing their transit service, more research is being devoted to this area.

In networks with many routes it is possible for two or more routes to share common sections, allowing the passenger to select which one to use. Waiting time, transfer time, and congestion all depend on the particular route chosen by a passenger, and thus the routes sharing common sections must each be considered explicity. The solution techniques developed for the traffic assignment problems cannot efficiently handle the transit route structure. A more accurate description of traveler behavior is needed, as well as alternative methods for finding shortest paths.

Chriqui and Robillard suggest a procedure for taking into account the frequency and speeds of competing routes (2). They develop an heuristic technique in a probabilistic framework which reflects the assumption that a passenger waiting at a stop will not let a vehicle go by and wait for a vehicle with a longer travel time. They concluded that the route selection process is sensitive to both the vehicle travel time and also the waiting times.

Le Clercq has developed an efficient algorithm for determining shortest paths through a routed network which considers, walking time, waiting time, travel time, frequencies, and transfer time (14). His procedure basically constructs a revised network which is then solved by a common shortest path algorithm. Although this procedure is advantageous in that it is efficient and the revised network is considered implicitly, conserving computer storage space, it is not capable of handling the required flow-dependent arc travel times. As will be discussed in Chapter IV, the concept underlying his procedure is applicable to variable travel times, and is used in this research effort.

There has not been much effort made to combine the assignment procedure for a network containing routes with the demand distribution phase that normally precedes in the sequential transportation planning process. From the literature review, it appears that such a model is needed, and that current mathematical programming techniques with today's computers are not yet sophisticated enough to solve such a model.

CHAPTER III

MODEL DESCRIPTION

A transportation planner faced with the task of designing an urban transit system must answer two basic questions concerning a possible system: 1) the amount of potential demand utilizing the system must be determined, and 2) given a certain number of travelers using the system, what paths will they follow? In most urban areas, the automobile is and will remain a major, if not the major, mode of transportation. The intention of transit planners, then, is to provide an alternative mode of travel that will serve those who cannot or wish not to commute by automobile, and that will relieve the typical congestion and energy waste associated with automobile use. For convenience, this second mode will be referred to as a bus system; however, a rail or street car system would perform the same way in the context of the model.

In order to compare the two travel modes and the different paths within the bus system, some decision criteria used by potential travelers must be determined and evaluated. These criteria can be reduced to a single, easily evaluated form, namely Perceived Total Travel Time from origin to destination. It is assumed that a fixed origin-destination demand will use one of the two possible modes. Thus, the number of travelers perferring one mode over another can be determined from a type of preference or demand curve based on the ratio of perceived total travel time on the system to the travel time by automobile (12).

The different paths available in the bus system are determined mainly by route structure. As passenger loads on the routes change, the perceived path travel times will change accordingly.

These concepts then can be used to reflect traveler behavior in the urban area. A further description of these ideas and their interrelationships follows.

Perceived Travel Time

In evaluating the possible ways to travel from an origin to a destination, a traveler makes a choice based on several criteria. Attributes such as actual travel time, waiting time at a pick-up point, proximity of bus stops, comfort of ride and many others are all considered in the decision process (11). In this model these attributes must be evaluated in order to choose paths within the bus system and also to choose between the automobile and the bus modes.

The most convenient way to evaluate these attributes is to define a common denominator to which all the attributes can be easily converted. Two common bases for evaluation are functions of dollars and of travel times. The dollar based denominator has been developed on a theoretical basis, and provides insights into the relationships of the various factors affecting the ridership on a transit system (16). However, the generalization to the dollar base is quite complicated and several assumptions concerning costs, interest rates, etc., make the conversions and comparisons too difficult to be practical at this time.

On the other hand, a function of time is a convenient way of combining the different factors. For example, Hoel (10) has developed a model which determines "total" travel time, considering walking, waiting, and movement times. Each of these times is given a different weighting to arrive at the final figure. The walking and waiting times are related to the transit system's proximity of stops and frequencies, respectively. Generally, waiting time is considered the most important of these, especially when the journey requires a transfer between routes.

The movement time is usually considered simply the actual time spent on the transit vehicle. However, the capability to combine and include other factors affecting passenger behavior does exist; specifically, the discomfort a passenger may experience during a journey definitely changes his perception of the actual time spent moving. This concept of a discomfort factor leads to the definition of "Perceived Travel Time".

A transit system can only have a certain number of seats available for passenger use. As more passengers board the system, some are forced to stand. Having to stand can be interpreted as a discomfort factor, thus affecting the passenger's perceived travel time. In the example presented earlier in Chapter I the travel time was perceived as double the actual vehicle travel time. The discomfort factor, however, is not to be interpreted as a simple step function in terms of the vehicle.

Realistically, the discomfort on a vehicle increases as the vehicle crowding increases. A standing passenger who boards a mildly loaded bus (all seats taken but few standees) will be more comfortable than one who boards a heavily loaded bus, since in the former case the passenger can select his standing spot and hand support and usually defend his territorial space.

One can then think of journeys with discomfort factors affecting perceived travel time in terms of network arcs with variable, flow dependent costs. As more flow is put on an arc the greater the cost for the next increment of flow assigned. The increasing costs of the flow will act as a penalty function, tending to limit the flows on the arcs in the network. This makes sense in terms of the bus system, as only a finite number of passengers can board a particular bus before the crowding effect will render the bus completely undesirable for additional passengers.

The perceived movement time on a bus can be described as a piecewise linear function. As long as there are seats available, the perceived time remains constant, approximately equal to the actual vehicle travel time. Then, once the number of passengers on the bus exceeds the seating capacity, the perceived time will increase as the ridership increases. Since the waiting time and walking time are not flow dependent, they can be included directly with the actual movement time of the bus, which can then be called the base travel time for the journey in the transit system, a constant. For simplicity in the design of this model, the increase in perceived travel time as demand increases beyond the seated capacity is assumed linear. This is not an essential or restrictive assumption as non-linear functions may be substituted for the discomfort factor with little effort. A diagram of the perceived travel time versus ridership between two stops is given in Figure 1. Note that the curve is constant until the seated capacity is exceeded.

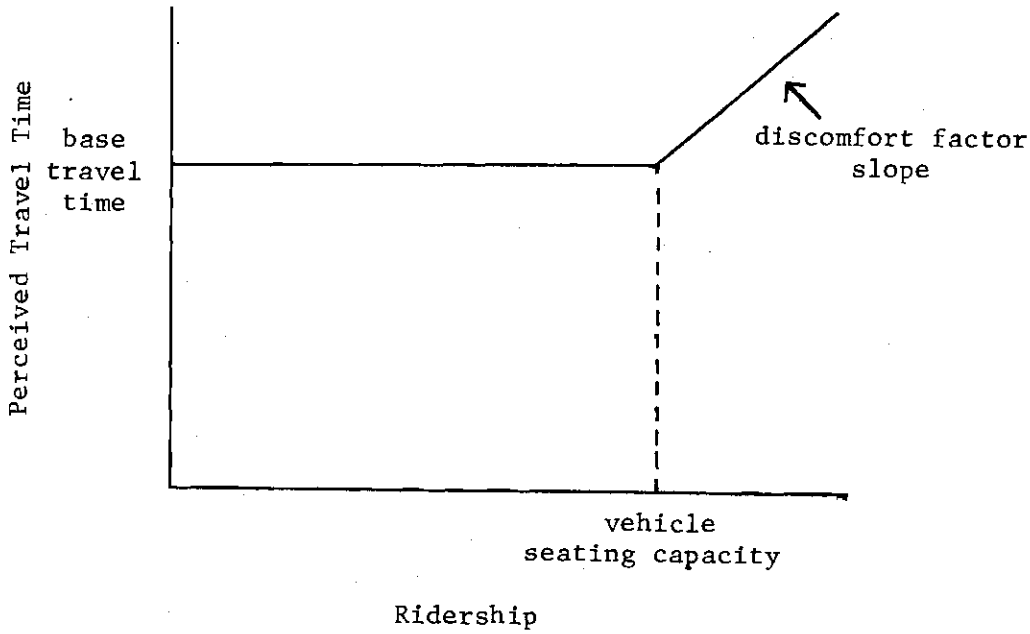

Figure 1. Perceived Travel Time Function

The effect on the perceived travel time of combining several such simple arcs into a longer journey is discussed under demand elasticity and in Appendix B.

The Demand Function

The availability of two modes of travel from an origin to a destination has already been discussed, and a criterion for discriminating between the modes has been established. Now will be presented the concept of how travelers use this criterion to choose between the modes; that is, how to use the traveler's perceived travel time to model actual traveler behavior.

If an ideal mass transit system could be designed that would reduce door-to-door travel time to zero, virtually everyone would want to use it; as the perceived total travel time between an origin and destination increases, fewer travelers would use the system. Eventually, when the perceived travel time becomes greater than any other alternative mode, almost all the travelers will leave the system (1). It is the distribution of the demand between these two extremes that will now be developed. The development of this distribution follows closely the arguments given by Aburto and others (1) for the general demand distribution case.

For this model, the base or zero-load travel time for the transit system will be defined in terms of the perceived travel time for a similar journey by automobile. Thus, whenever the perceived travel time for a trip by bus equals that by auto, a certain maximum number of travelers for that origin-destination (O-D) pair will choose the bus system. This maximum number of travelers would usually be a fraction of the total number of travelers for the O-D pair, since some travelers will still prefer the automobile. On the other hand, once the trip by bus becomes too long compared to the auto time, all travelers will choose the auto mode. This behavior for a particular traveler can be described in graphical form as follows. Let auto be the corresponding travel time by car. The point t* is the point at which the perceived travel time is so much greater than the auto travel time that the passenger will leave the system.

The perceived travel time for a given journey will vary from one traveler to another, as their evaluations of the discomfort factor may not be the same. For example, an older traveler may strongly

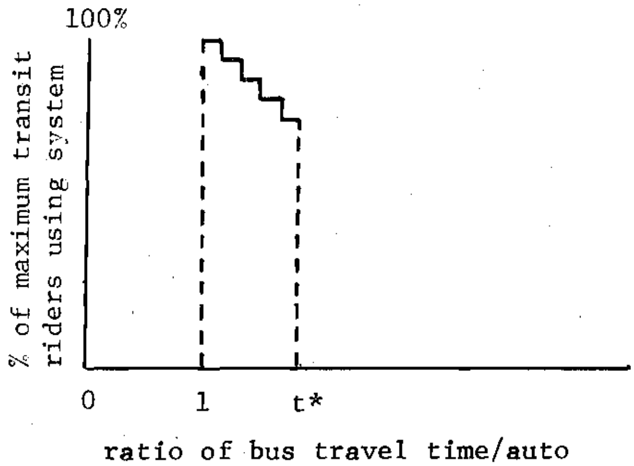

Figure 2. Traveler Behavior

object to standing during a journey whereas a younger passenger may not discriminate between sitting or standing. In this case, the older traveler may weight the time spent standing 10 times as much as seated time, greatly increasing his perceived travel time in comparison to other travelers.

The value t* then, will vary among the total potential ridership. It is reasonable to assume that the values of t* will correspond to identically distributed random variables. It can be shown that the actual distribution of demand for large populations can be approximated by a continuous distribution as shown in Figure 3 (4).

All that is left to describe completely the passengers' behavior then, is to determine a function which will represent the demand curve with respect to travel time. To provide better insight into the capabilities and interrelationships within the model, and also to relieve unnecessary computational difficulties, a simple approximation to the demand curve is preferred over a more exact and

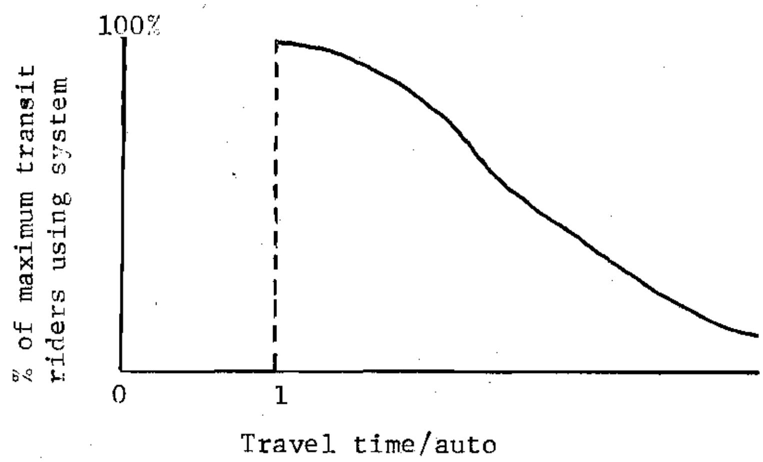

Figure 3. Traveler Demand Curve

at the same time more complicated function.

In developing the function, three aspects of the model must be considered. First, there exists a base travel time (auto travel time) for a path at which a certain maximum number of travelers for that O-D pair will use the transit system. Second, each bus has associated with it a seated capacity, which when exceeded by the number of travelers on the bus, causes the perceived travel time for standees to increase. Third, there exists some travel time for a path on the system above which no travelers will use the system, although this last restriction could easily be relaxed.

Considering these aspects and the desire for a simple function, a linear function is suggested. For perceived travel time on a bus for a journey which is equal to the corresponding auto times the function gives 100% of the maximum available riders. (Bus travel times less than auto travel times are not considered, although they could easily be incorporated.) Then, as travel times increase beyond the base time a linear decline in the number of passengers is expected.

This linearity is appealing when considered in conjunction with the linear perceived travel time, and it also guarantees the existence of an upper bound on travel time which must be considered. The following diagram illustrates how the linear function approximates the smooth demand function.

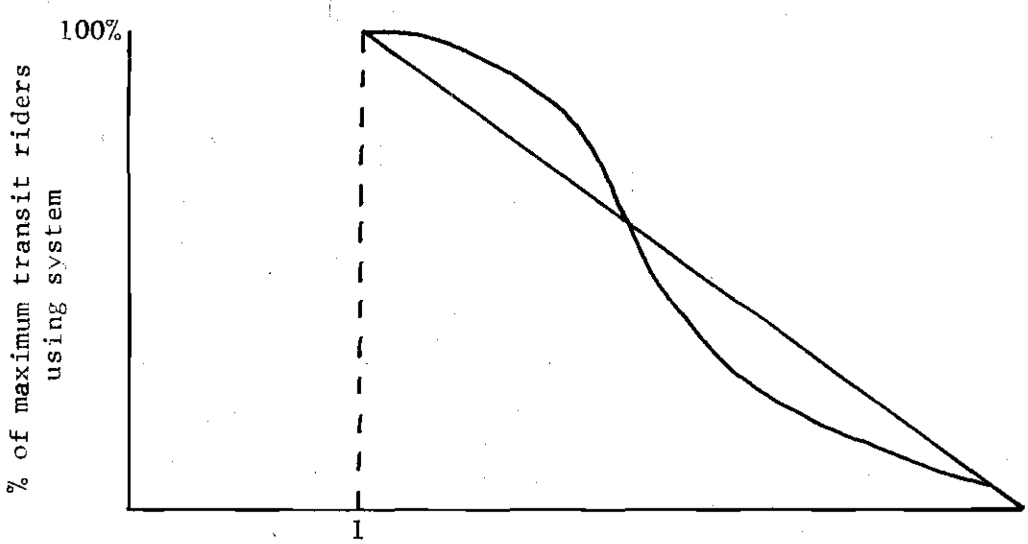

Figure 4. Linearized Traveler Demand Curve

The traveler behavior then can be described by two variables; the auto time for a particular journey and the slope or elasticity of the demand curve. Then, given the perceived travel time, the percent of maximum demand which will utilize the system can be determined.

Route Structure

In general, mass transit system vehicles follow a particular path or route repeatedly. The few examples of personal rapid transit (PRT) systems, in which vehicles do not routinely follow a prescribed path, are not well described by this model. The importance of the vehicle route structure in the transit system is realized when its effects on the perceived travel time are considered.

In early morning rush periods, the flow of travelers is characterized as largely originating in the suburbs, and terminating in the central downtown area (12). Thus, as buses approach the central urban area from the outer edges the number of passengers will increase. It is reasonable to assume then, that as a bus starts from the end-point of a route, at the outer edge of the urban area, the load on the bus would, in general, be the least. Potential riders who might wish to board the buses at these extreme points then have the best chance of getting on the bus, and also of getting a seat, keeping their perceived travel time relatively close to the actual riding time.

Once the passengers are on a bus, they effectively ignore intermediate stops along their journey. This is a different situation than say, a traffic assignment model, where at any node no traveler entering the node has a priority on arcs leaving the node, allowing for travelers to be forced off the system by congestion at intermediate stops. Obviously, once a traveler is on the bus, and particularly if he has a seat, he is not forced to compete for space at every stop with newly boarding passengers. This is particularly true for a seated passenger, and to a lesser degree for a standing passenger. The latter can usually select his standing space, maintain his handhold, and defend his territorial space. Only at the initial boarding point or at a transfer point to change routes must a passenger compete for space on a bus and for a seat.

Thus, the discomfort factors which are used to compute perceived travel time are based largely on the bus loadings at passenger origin and transfer points, or each time a passenger boards a route. Passengers who undergo a transfer have an additional waiting time (and also a transfer time) included in their perceived path total travel time. The result of this is that the bus load on the second route has a lesser influence on the perceived total travel time. In the inward loading procedure, the bus load on the second, usually outward bound, route is then estimated and used along with the known bus load on the first, usually inward bound, route.

Equilibrium Model Formulation

The descriptive model has now been developed considering the demand and travel time functions and the route structure. In the following paragraphs the problem is formulated as an equilibrium model.

First, denote the perceived travel time between origin a and destination b as t^{ab}. The demand function between a and b, then can be expressed as a function of this travel time by:

$$x^{ab} = x^{ab}(t^{ab})$$

and the inverse of the demand function by:

$$t^{ab} = g^{ab}(x^{ab}) \ .$$

The perceived travel time on a route m connecting nodes i and j by arcs $x_{mik}, x_{mk\ell}, \ldots, x_{m\ell j}$ is a function of the flow on the first arc traversed:

$$t_{mij} = t_{mij}(x_{mik}) \ .$$

The equilibrium sought, then, is a vector of flows which satisfies the functions for the demands and perceived travel times on the transit system. For automobile traffic models one can show (18) that finding the equilibrium flow is equivalent to a minimization problem where the objective function contains integrals of g^{ab} and t_{ij}, where the latter is the flow-dependent travel time on arc (ij):

$$t_{ij} = t_{ij}(x_{ij}) \; .$$

CHAPTER IV

SOLUTION PROCEDURE

The solution procedure developed for the model is an heuristic procedure composed of three separate iterative algorithms. The first algorithm begins with no commodities loaded on the transit system and assigns paths and effective demands for each commodity, considering all commodities with identical origins simultaneously. For convenience, a commodity is defined as a group of travelers corresponding to an O-D pair. Each commodity is handled only once to obtain the initial solution. Second, the commodities are again considered by nodal groups; their paths and demands are recalculated based on the status of the system yielded by the initial solution. In the third and final algorithm, the paths of the commodities remain fixed and the demands are recalculated based on the perceived travel times on the paths.

The Choice of Heuristic vs. Exact Procedure

In solving assignment problems both heuristic and exact procedures are sometimes used.

In Chapter II, the advantages and drawbacks of both methods were discussed. The two most important features of the model which affect the solution procedure selection are that:

1) the model does consider a network with route structure and,

2) elastic demand functions must be handled.

There are no practical exact procedures which can handle the route structures inherent to the transit system; the problem is simply too large. Add to this the elastic demands and such a procedure is totally unrealistic. For this reason an heuristic procedure approach has been adopted for the solution algorithm. The only major storage needed is a trip matrix, the elimination of which is considered in Chapter VI.

The Initial Pass Algorithm

The initial pass algorithm is basically an iterative procedure which considers all commodities originating at a single node, determines their paths and demands, and moves on to the next node until all commodities have been considered. The procedure starts with a completely unloaded system, that is, no commodities have been assigned to a path. The desired result from this initial pass is an initial solution, or loading of commodities on the system, which will serve as a starting solution for the second pass algorithm. Since the system is initially empty, the order in which the commodities are considered is critical. This is a result of the route structure and its effects on congestion, since a rider boarding a bus at a point near the route origin will affect the chances of a rider further along the route getting a seat, whereas the reverse situation will not occur. Also, at each stop along a route all commodities boarding the bus compete simultaneously for seats and standing room.

The latter problem is handled by actually considering all commodities originating at the same node simultaneously. This reduces the problem of having to order the commodities and shifts the emphasis to ordering the nodes. The most appealing ordering scheme is based on the route structure and the concept that the perceived travel time of passengers boarding a vehicle is affected only by other passengers on the vehicle who board nearer the route origin, and not passengers who board nearer the destination.

This observation suggests that the first node to be considered should be at an origin of a route. Then one should consider all such nodes at the route origins and progress through all the nodes, always considering the available node nearest a route origin. It is possible to proceed in this manner until all nodes have been considered. As was discussed in Chapter III, the route origins, in general, are the stops furthest from the central downtown district. So an equivalent ordering scheme is to consider the nodes furthest from the central district first and proceed inward, toward downtown. This scheme would tend to minimize the effects of determining perceived travel times on an empty or near empty system.

Once the order to consider the nodes is established, the next step is to determine the path, path length and effective demand for each commodity at a node. Ideally, this step could be performed with an exact procedure, computing all three attributes at one time. Due to the non-linearity in the travel time and demand curves resulting from the interaction of different commodities competing for seats and standing room even from the same origin, the calculations are too complex and complicated to make an analytical solution tractable.

Suppose we have a network of two nodes, one arc and one commodity. The base travel time for the commodity with potential demand of 10 passengers is 6 minutes. The perceived seated travel time by bus is 8 minutes and there are six seats available. The discomfort slope is +1 and the demand slope is -1. This case is shown graphically as follows: (Appendix B describes how to rotate and expand the demand curve axis to help visualize the two curves easily).

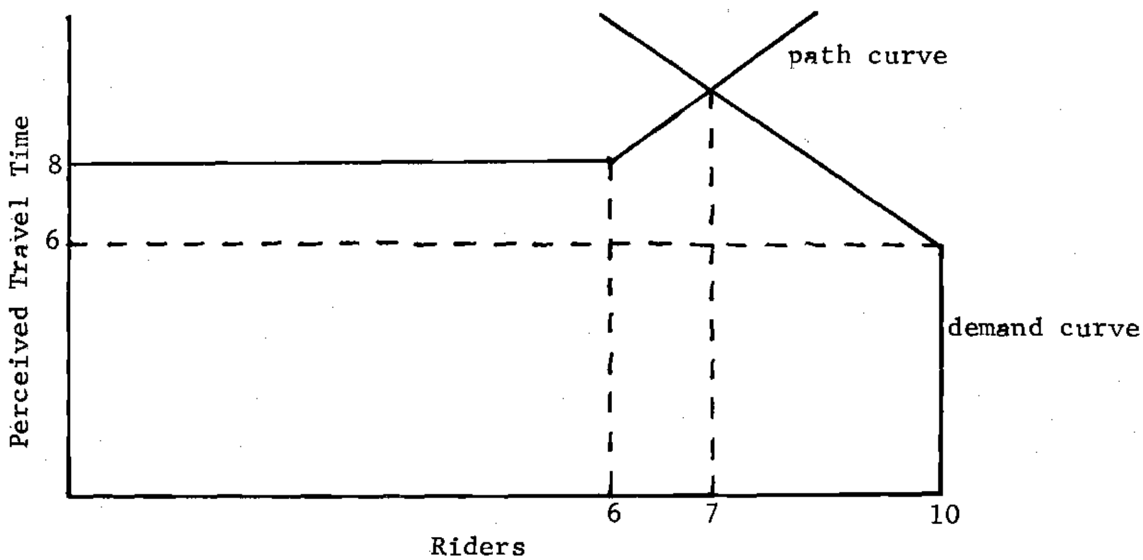

Figure 5. Single Commodity - Single Arc Path Curves

The solution to this simple problem is where the two curves meet; in this case a demand of 7 riders and a travel time of 9 minutes. This is an equilibrium point, that is, with seven riders on the bus the perceived travel time is nine minutes, and with a perceived travel time of nine minutes, seven people board the bus.

Appendix B extends this simple problem into a single commodity multi-arc multi-route problem and shows how when more than one commodity is considered, the changes in perceived travel times for the several different paths, which share some but not all arcs in their paths, become unmanageable. For this reason the initial pass algorithm makes use of an approximation procedure to determine the three commodity attributes.

Determining Shortest Paths, Demand, and Path Lengths

The approximation procedure consists of two parts. First, the shortest path for each of the commodities is found based on the previous nodal assignments and then second, the demand and path lengths are estimated.

The shortest paths are found by constructing a shortest path tree from the origin node to all other nodes in the network using a modified labeling technique. Since the network contains a route structure, common tree building algorithms are not very efficient, or can not always produce shortest path trees. For example consider the simple network below, where a waiting cost of ten minutes is incurred for a transfer between routes.

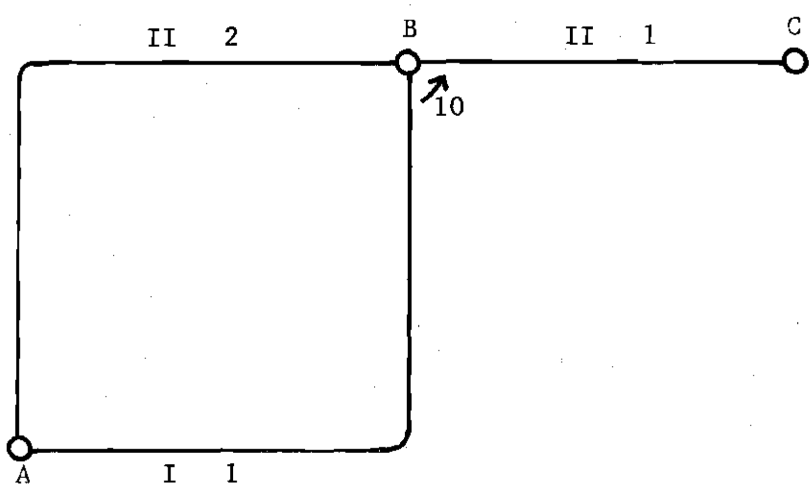

Figure 6. Network with Route Structure

The shortest path from A to B is via route I with a travel time of one minute. The shortest path from A to C though is via route II with a travel time of three minutes. The usual shortest path tree would not have permitted this path, however, as it means node B is reached twice from node A.

Le Clercq has developed a procedure which builds a revised network that handles the routes and yields true shortest paths (14). Appendix A explains how his method is adapted for use in this solution procedure.

Once the paths have been determined, the demand of each commodity and the corresponding path length must be estimated. This is similar to a classical economics problem where the equilibrium between supply and demand is desired. In these equilibrium cases the desired point is usually obtained by a spiralling procedure, where supply and demand are iteratively defined in terms of each other until they converge within a given range of the equilibrium. In the travel time - demand case however, such a spiralling procedure pursued for several iterations for all commodities would be very expensive in terms of execution time.

The algorithm performs only two such iterations and then an estimate of demand and path length is made based on these few iterations. It is hoped that the execution time saved by making perhaps a poorer estimate can be put to better use in the later algorithms to yield a better final solution. The approximation procedure for several commodities at a single node is outlined below:

1. Store the shortest path length for each commodity, in an array (say in PTHLA),

2. Using the appropriate path length (PTHLA) determine the number

of riders for each commodity. Store these values in an array (say DEMA),

3. Determine new arc travel times and subsequent new path lengths _if_ all the demand found in 2) (DEMA) from all the commodities is assigned on the shortest paths. Store these new path lengths in an array (say PTHLB),

4. Repeated step 2) using PTHLB and storing demands in DEMB,

5. Take a weighted average of DEMA and DEMB to estimate the final demand assigned,

6. Load this demand on the shortest paths.

The demand calculated in step two represents the maximum number of riders from each commodity ever expected to board a vehicle, much like an upper bound. Since this demand is an upper bound, then if all of it were loaded the resulting path length, by the nature of the discomfort curve, would be at an expected maximum, or upper bound. If this path length is an upper bound, then, by similar arguments the demand found in step 4 can be thought of as a lower bound. Thus the desired demand lies between the two calculated demands for each commodity. Notice that in step 3 all commodities are loaded and then the path lengths are updated; this tends to account for the commodity interaction.

The estimate of the final demand depends heavily on the weighting scheme used for averaging the two demands; however, the amount of available information to use as weights is limited. The weights chosen are the path lengths which were used to calculate the respective demands. There is no rigorous analytical explanation as to why such weights should be used, but considering the available information it appears to account better for the available seats on the vehicle than say an equal weighting. Precisely, the weighting for each commodity is performed as below:

$$\text{DEMAND} = \frac{\text{PTHLA*DEMA+PTHLB*DEMB}}{\text{PTHLA+PTHLB}}$$

Once this averaging process is complete the algorithm moves on to perform the same procedure for the next node. When all the nodes have been considered and the initial solution obtained, the solution procedure begins the second algorithm.

Second Pass Algorithm

In this second pass a criterion is defined which determines if the commodities at a particular node should be reassigned, changing their paths and/or demands as the situation requires. The purpose of the pass is to locate any commodities which in the initial pass were improperly assigned, perhaps due to a faulty assumption at a transfer point when the system was lightly loaded, or the nature of the route structure misled the node ordering scheme. The criterion used is determined as follows:

1. Given the demand from all commodities in the previous solution, update all path lengths,

2. Given the new path lengths, determine and store the corresponding new demand for each commodity,

3. Determine the absolute difference between the assigned and new demand and call it the discrepancy. The discrepancy, or the

difference between assigned demand and "desired" demand is the defined criterion. It can be interpreted as an indicator of how close the system is to an equilibrium state.

The algorithm then picks the node with the commodity originating from it with the largest sum of discrepancy and performs the same operations as a node considered in the first pass. It is expensive in terms of execution time for the algorithm to calculate discrepancies. Therefore, depending on problem size, more than one node may be reconsidered between calculations of discrepancies with the order based on the of the relative discrepancies.

The maximum number of nodes considered in this pass will vary depending on problem size. After so many reassignments the second pass may reach a point where the execution time may be better spent in the third pass. This is especially true when the cycling problem is encountered.

Cycling in the Second Pass

Due to the nature of the arc travel time curves and the method for estimating demands at nodes, the cyclic repetition of identical demands used in determining discrepancies sometimes occurs. This repetion is most easily explained with an example. Suppose after the approximation procedure that the assigned demand from a node on an arc does not require the total number of available seats. This means that the path lengths before and after assigning the demand did not change. Now if the upper bound was greater than the number of seats, cycling will occur.

What happens is the demand calculated in the second pass to compare with the assigned demand in computing the discrepancy will always equal the upper bound demand calculated in the approximation procedure. This is because the assigned demand did not use up all the seats so the perceived travel time did not change. The diagram below illustrates on one demand curve what takes place.

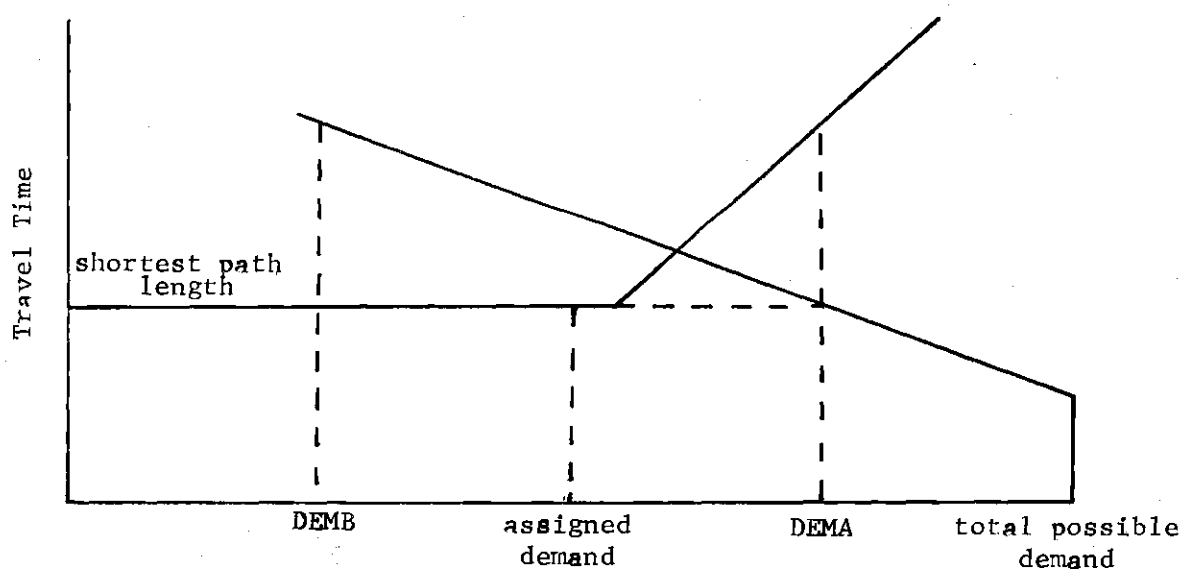

Figure 7. Example of Cycling in Second Pass

Obviously, this is a flaw in the approximation procedure which results from the linearity of the curves. However, when the purpose of the second pass is considered, cycling is not a problem which must be dealt with. Each commodity must only be considered once in this pass since the purpose is to reassign commodities based on a loaded rather than empty system. So once a commodity is considered in the second pass it is flagged and ignored until the third pass begins. This flagging procedure eliminates the possibility of a commodity cycling more than one time.

Third Pass Algorithm

The third pass algorithm assumes all commodities have been assigned to their shortest paths and all that must be done is adjust the magnitudes of the assigned demand. The same criterion as in the second pass is used to determine which demands must be altered. The basic steps are:

1. Calculate discrepancy as in pass 2 and pick commodity with largest discrepancy,

2. Perform the following steps for all commodities whose initial arcs are the same as that of the commodity in 1).

NOTE: CALL CURRENT ASSIGNED DEMANDS = DEMA

CALL NEW DEMANDS USED IN DISCREPANCIES = DEMB.

calculate new assigned demand by following formula:

$$= \frac{\text{DEMA+DEMB}}{2} \quad \text{if} \begin{cases} \text{DEMA and DEMB} > \text{No. seats} \\ \text{or} \\ \text{DEMA and DEMB} < \text{No. seats} \end{cases}$$

$$= \text{DEMA-1/2(DEMA that is standing)} \quad \text{if} \begin{cases} \text{DEMA} > \text{No. seats} \\ \text{and} \\ \text{DEMB} < \text{No. seats} \end{cases}$$

$$= \text{DEMB-1/2(DEMB that is standing)} \quad \text{if} \begin{cases} \text{DEMA} < \text{No. seats} \\ \text{and} \\ \text{DEMB} > \text{No. seats} \end{cases}$$

3. GOTO 1)

The two main differences in this algorithm compared to the previous two is that the paths remain constant and that commodities are grouped by initial arcs, not origin nodes. This algorithm is begun under the assumption that the final solution from the second pass is close enough to the equilibrium that no more path changes are necessary.

The function in step 2 for recalculating the assigned demand is basically a bisecting search designed to eliminate any cycling as it appeared in pass two. If both demands used in determining the discrepancy agree to allow 1) only riders with seats or 2) standees and sitting passengers, then a simple bisection between the demands is made, as cycling will not occur. If one demand allows strictly riding passengers and the other standees then the function calculates a new demand which avoids cycling. This demand is determined by taking the greater of the two demands (i.e. the demand allowing standees) and cuts down the number of standees in half, still leaving all the seated passengers on the system. This calculation effectively bisects the standing riders rather the total number of riders.

Once the discrepancy of the demands of the commodities is reduced below some set minimum value or a maximum number of demand calculations have been made, the third pass terminates. The algorithm can then return to the second pass and find new shortest paths based on the new flows. The procedure recycles between the second and third passes until some stopping criterion is met.

Since there is no obvious way to determine how close the solution is to the equilibrium at any point in the algorithm, a stopping criterion based on some definition of solution quality is not practical. For this reason the process is stopped when the amount of execution time expended exceeds some preset time limit. There are convenient checking points in both the second and third passes to terminate the process and print the final solution.

In summary, the solution procedure begins with an empty system and sequentially makes an initial assignment of flow to the network. The algorithm then enters a second phase which reassigns the flow if warranted by congestion effects. The third phase then adjusts the flow on all paths without changing paths. The procedure cycles between the second and third pass until the specified execution time in spent. The flowchart in Figure 8 represents this procedure.

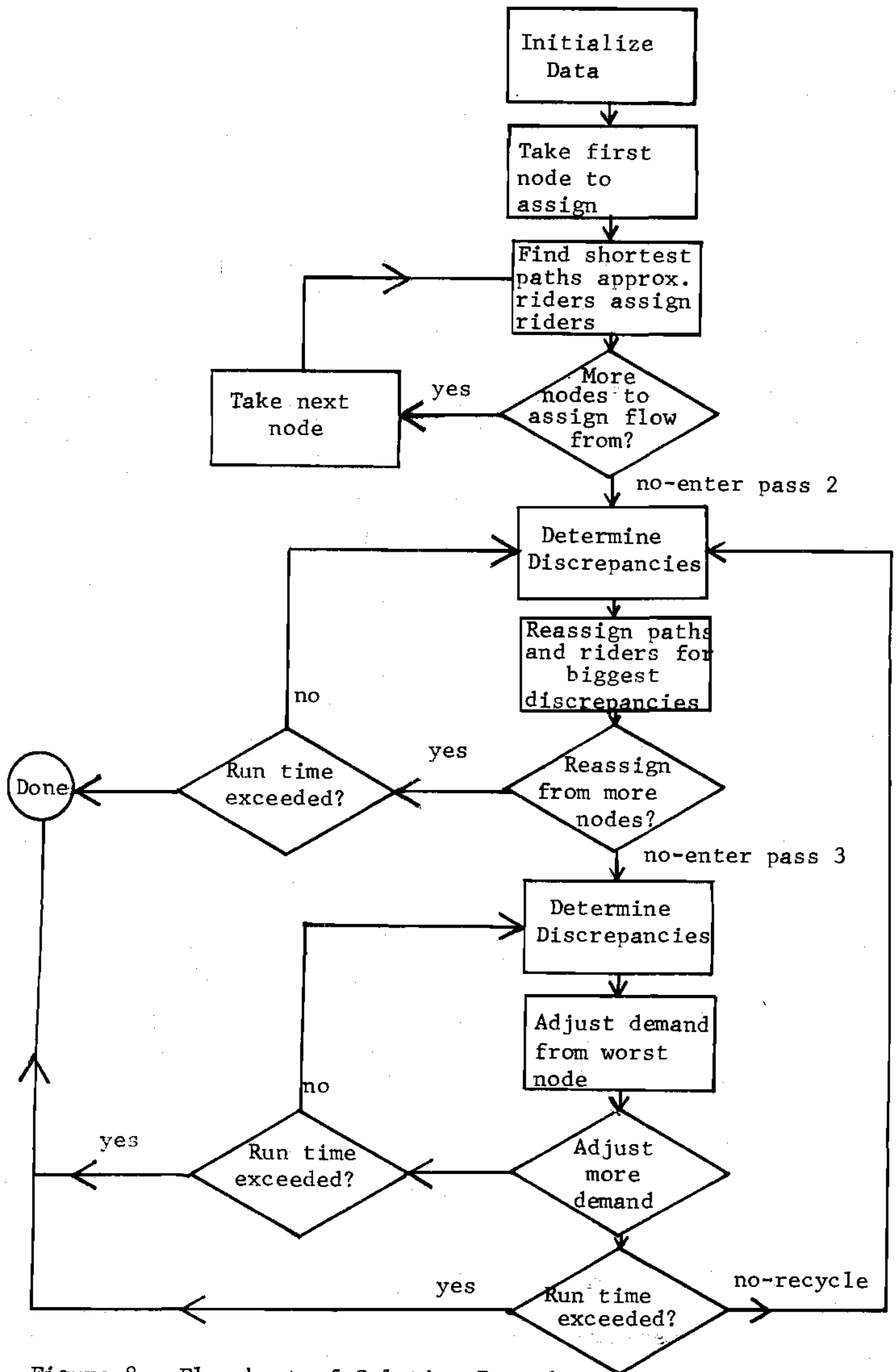

Figure 8. Flowchart of Solution Procedure

CHAPTER V

COMPUTATIONAL EXPERIENCE

The solution procedure for the model was coded in FORTRAN IV for use on the CDC Cyber 74 at the Georgia Institute of Technology. Some test runs were then made using randomly generated problems to gain some experience with the algorithm. This chapter will describe these tests, including the measurement of solution quality, the input data, and the results of the tests.

Measurement of Solution Quality

The solution procedure is an heuristic algorithm with no guarantee of convergence. To evaluate solution quality, then, two criteria were considered: first, the execution time required to reach a given level of convergence; and second, the level of convergence achieved in a given execution time. Now, measuring the level of convergence, even to a local solution, is time consuming, requiring about 10 seconds for a typical 30-node network. In addition, little information was known about what convergence level might be achieved. For these reasons, the second criterion was adopted, namely, the level of convergence achieved in a given execution time.

Since the true solution, or solutions, is unknown, the characteristic to be measured is the total amount of change in demand if each commodity is allowed to change paths (and presumably path lengths as a result) individually, with all other commodities remaining as assigned in the final solution. This change is best described by presenting the procedure by which it is determined, after the final solution is obtained:

1) Find new shortest paths and corresponding path travel times for all commodities, based on the final solution.

2) Calculate the demand associated with each commodity based on the new path travel times.

3) Compare the demand in the final solution to the demand computed in step 2 to determine the amount of change.

In order to compare runs which have different input data, the change is represented as the percent change in the total potential demand. Although the change calculated in this manner should not be used as an indicator of how close a solution is to the equilibrium, it is appropriate for use to compare the quality of two solutions.

Input Data

The networks used as test problems were generated randomly using a computer routine provided by G.P. Sharp. The nodes, arcs and O-D pairs and their attributes were the outputs used to create the networks. Unless otherwise noted, the test problems all used the same network, with modifications to the route structure distinguishing the problems from one another. Also, it should be noted that the commodities reflected a morning rush period, with the heavier flows generally toward the center of the network.

To limit the difficulties arising from random error, replications of each problem were generated by randomly varying the amount of potential demand for each commodity. This was accomplished by using the following formula:

POTENTIAL DEMAND = BASE POTENTIAL DEMAND*(1+RANDOM)

where RANDOM is a random number from a Uniform (-.25,.25) distribution.

Three basic test problems were created using the methods described above. They each had 30 nodes, 340 O-D pairs and 10, 8, or 5 routes. The slopes of the demand and travel time curves were held constant during the testing except during noted trials. The demand slope was set at -.5. This meant that if the journey by bus took twice as long as by auto, the number of riders would be one half of the potential transit demand, a reasonable assumption for these purposes. The travel time slope used was chosen, such that a journey with a bus having as many standees as seats, had a perceived travel time twice the magnitude of the actual vehicle time.

Test Results

The main body of the tests was centered around three parameters of the solution procedure: the order of the nodes in the initial pass, the number of nodes in each iteration in the second pass, and the number of routes in the network. Other test runs were made hoping to yield some insight into areas such as solution time for larger networks, and sensitivity to changes in the perceived travel time slope.

Factorial Design

The tests on three parameters of the procedure were made using a 3^3 factorial design. Each cell had two replications, resulting from changes in potential demands. A certain amount of initialization is required at the beginning of each run to accomodate different data sets in a convenient manner. The execution time was set at 70 seconds beyond the initial set-up time for all runs. These results appear in Table 1.

An analysis of variance on the results shows the node ordering to be significant at the 1-percent level, the number of routes at the 5-percent level, and the node ordering interacting with the number of routes at the 5-percent level. Table 2 presents the details of the ANOVA. The first two results confirm expectations about the behavior of the solution procedure.

The node order analysis reflects the assumptions concerning the initial pass with the system unloaded. Beginning with the nodes near the route origins, the solution after 70 seconds was much more stable (mean = 4.4%) than with random ordering (mean = 6.0%), and loading from the center out was the worst case (mean = 6.6%).

These results are further verified by considering the discrepancy immediately after the initial loading. Loading from the route origins inward the initial solution was again much more stable (mean = 6.8%), than either the solution arrived at with random ordering (mean = 12.0%) or center out loading (mean = 16.3%). Tables 3 and 4 present the discrepancies after the initial loading and the details of the ANOVA respectively. With this data the node ordering is significant at the .1-percent level.

The significance of the route structure tends to reflect the assumptions made by transit planners in the past: as the number of paths becomes small in the transit system, the assignment procedure is much more stable.

Table 1. Percent of Potential Demand to be Changed After 70 Seconds Execution Time

	Number of Routes								
	10 Routes			8 Routes			5 Routes		
	Number Nodes/Pass								
Initial Node Ordering:	3	5	10	3	5	10	3	5	10
Outside Inward	5 5	4 4	5 4	5 5	5 4	4 4	5 5	4 4	5 5
Inside Outward	8 5	6 5	8 5	5 4	5 4	6 5	7 4	9 6	8 5
Random	9 5	5 4	7 4	6 5	6 5	7 5	5 4	6 5	6 5

Table 2. ANOVA for Factorial Design

Source	df	SS	MS
Node order (A)	2	46.8	23.4**
# iterations per pass (B)	2	3.4	1.7
# Routes (C)	2	6.7	3.4*
AxB interaction	4	2.1	.52
AxC interaction	4	15.1	3.8*
BxC interaction	4	5.2	1.3
AxBxC interaction	8	7.0	.88
Error	27	19.0	.70
TOTALS	53		

*One asterisk indicates significance at the 5-percent level; two, at the 1-percent level.

Table 3. Percent of Potential Demand to be Changed After the Initial Pass

	Number of Routes		
Node Ordering:	10	8	5
Outside Inward	7,6	9,9	8,8
Inside Outward	7,7	16,16	12,12
Random	7,7	24,24	16,16

Table 4. ANOVA for Percent of Potential Demand to be Changed After the Initial Pass

Source	df	SS	MS
Node Ordering (A)	2	271.4	135.7*
No. Routes (B)	2	184.1	92.1
AXB Interaction	4	105.6	26.4
Error	9	.5	.006
TOTALS	17	561.611	

* Indicates significance at the .1% level.

The interaction between the route structure and node-ordering is difficult to explain. Figure 9 shows the plotting of cell totals vs. the number of routes for the three node loading orders. The most obvious interaction results from the center outward node order and the eight routes.

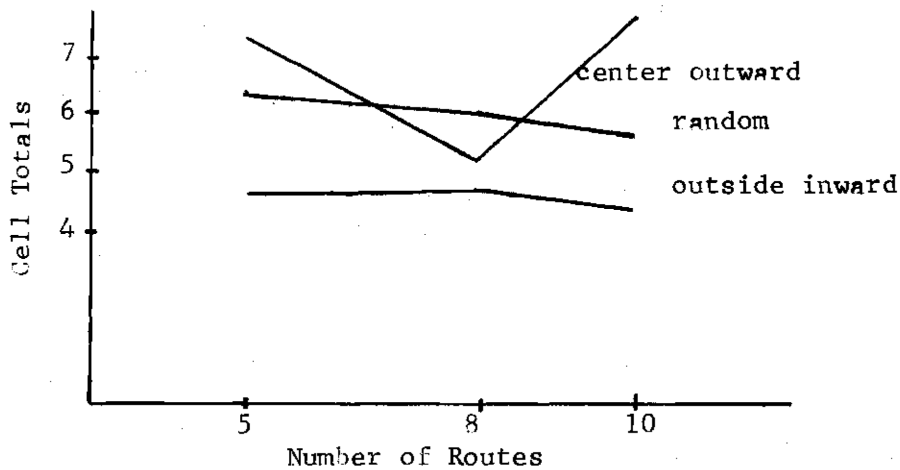

Figure 9. Cell Totals vs. the Number of Routes for the Three Node Ordering Schemes

A likely possibility is the nature of this particular network, causing severe non-equilibrium near the center when the initial pass was ordered this way, which has hard to correct when limited by the few routes.

Miscellaneous Test Runs

The first miscellaneous test problem used the same network as the general runs with 10 routes, 4 nodes per iteration in the second pass and outside inward initial node ordering. The execution time limit was set to 200 seconds, however, to determine if the solution was converging. It turned out that after the solution reached the point where the other runs were terminated, it did not significantly change thereafter. The procedure cycled through the second and third passes about 25 times with no appreciable improvements made.

This stabilization is the result of the poor discrepancy criterion used in pass two. Once the solution gets within a certain range of the equilibrium, the demands become very stable, thus the second pass does not detect the need to find new shortest paths, as found in the analysis of the final solution. Either a better criterion is needed, or, perhaps after so many cycles, new trees should be built for all commodities. The latter may be feasible as more efficient tree building algorithms are incorporated in the algorithm.

Another type problem tested was similar to the general problems, but with a much steeper slope for the travel time curves. The slope tested was on the order of 100 times greater than the original slope, which caused a major discontinuity at the point where the passengers were forced to stand rather than sit. The results were very poor: 41% of demand after 70 seconds execution time.

The discontinuity caused the approximation procedure in the initial pass and the second pass to be very inaccurate. The path length weighting factors tended to put the demands near zero for each approximation. The third pass then, was never able to converge toward an equilibrium as it was always forced to correct the bad approximations. If such steep curves are required, then a new approximation procedure is essential.

The third miscellaneous test problem run again used the same general network, only with a potential demand matrix which was much closer to being symetric rather than center oriented. This situation would model, say, the mid-day demand, as opposed to the morning rush period. The results were poor (28% of demand) when using the outside-inward initial node loading order.

The solution improved, however, with the use of the inside-outward ordering (8% of demand), although still not as stable as the first group of tests. This may be an indication that the assumptions made for the morning rush are not robust when the more general case is considered.

One characteristic which could explain much of the good behavior of the center oriented model is the convenient procedure for ordering commodities by nodes initially. When the opposite extreme occurs, that is, evening rush, serious problems could arise with the commodity ordering from a single source. In this case a procedure for allowing commodities to simultaneously "swap" routes may be needed.

The final test run was made with a much larger network on the order of seventy nodes, 1250 commodities and 23 routes. The execution was terminated after 2500 seconds. This time appeared to be excessive as the same stabilization and lack of convergence that was apparent in the first miscellaneous run again affected the large problem. A nearly identical final solution could have been achieved within about 1000-1200 seconds. The time to build shortest path trees increased by about 10 times, but the time to adjust flows in pass three increased linearly with the number of commodities.

These results point out the dependence of the efficiency of the algorithm on how quick shortest path trees can be built. If a faster tree building procedure, such as a modification of Le Clercq's procedure, was utilized then large problems should not require completely unrealistic solution times.

Table 5. Summary of Miscellaneous Test Runs

# Nodes	# Commodities	# Routes	Travel Time Slope	Node Order	Demand Matrix	Execution Time	Discrepancy After Pass One	Final Discrepancy
30	340	10	.02	Outside Inward	Morning Rush	200 Secs	6%	4%
30	340	10	2.0	Outside Inward	Morning Rush	70 Secs	30%	41%
30	340	10	.02	Outside Inward	Full	70 Secs	28%	-
30	340	10	.02	Inside Outward	Full	70 Secs	8%	-
70	1257	23	.02	Outside Inward	Morning Rush	2500 Secs	3% *	3%
70	1257	23	.02	Random	Morning Rush	2500 Secs	8% *	4%

*Time required for initial pass was apporixmately 200 seconds.

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The objectives for this research were to: 1) construct a descriptive model of the combined distribution-assignment process for an urban transit system, 2) develop a solution algorithm for the model and, 3) gain some computational experience with the algorithm. In Chapter III the model was described; specific characteristics considered were demand elasticity, perceived travel time and route structure. The solution procedure developed and tested in Chapters IV and V showed an equilibrium of the model could be approached successfully with an heuristic algorithm.

The most important result of this research effort is the realization that a distribution-assignment model which considers demand elasticity and route structure is both practical and feasible. Such a model would be useful to urban transit planners who are still solving the demand-assignment problem sequentially, just when combined models are starting to be used in the urban traffic planning.

The solution procedure was never able to reach an equilibrium for a network. It was, however, able to converge until the improvement criterion became insensitive to the current solution. The solution procedure was successful enough to warrant further investigation into possible modifications for improving its performance.

The scheme developed for ordering commodities and initially loading from an empty system was one of the most significant aspects of the solution procedure. The quality of the initial solutions achieved in the test problems (after pass one) indicates the success experienced with the procedure for the morning rush model.

Recommendations

After making basic assumptions about the model and working with these assumptions, specific recommendations can be made based on difficulties and successes with finding a solution procedure and trying it on test problems. The recommendations are:

1) Remove the linearity constraint on the demand and travel time curves. This linearity assumption was made very early in the research effort as a simplifying advantage, but resulted in creating great difficulties in the solution procedure.

2) A better way to determine when new shortest paths should be calculated is needed. At one point during this research, alternative path lengths were found along with each shortest path to be used as an upper bound criterion. Perhaps further investigation in this area would provide such a better criterion for new paths in the second pass.

3) Although all the demand and travel time curves were considered identical in the model, it would not be difficult to define a different curve for each group of travelers or type of vehicle based on, for instance, social class or vehicle size.

4) Remove the all or nothing type assignment restriction, perhaps coupled with deleting the shortest path matrix. Just as in traffic

assignment procedures, once the number of different paths through a transit network multiply, it is no longer correct to assume all riders choose the same paths. Elimination of the shortest path matrix would save close to half the storage requirements of the computer routine, allowing much larger problems to be handled.

APPENDIX A

SHORTEST PATH ALGORITHM

In Chapter IV the need for finding several shortest paths from a given node was defined. Two important functions of the shortest path algorithm is that paths must be stored in a convenient manner to be frequently retraced and the route structure must be considered. Le Clercq (14) has developed such an algorithm which finds shortest paths through networks containing routes and stores the paths in an easily accessible matrix form. His procedure, however, does not consider flow dependent travel times which the model solution requires. For this reason it was decided that conversion of Le Clercq's algorithm to handle the variable travel times would be difficult, and instead the idea behind his algorithm would be coupled with a basic tree building algorithm to yield a procedure which will handle both variable travel times and capacities.

In order to consider the route structure in the network, Le Clercq redefines arcs as being "a directed movement possibility in the network between two stops which could be performed without transfer". From this definition he can implicitly create a revised network in which each node served by a given route in the original network is connected with single arcs to all other nodes served by the route.

The example given in Chapter IV is repeated on the following page with the revised network shown beside it.

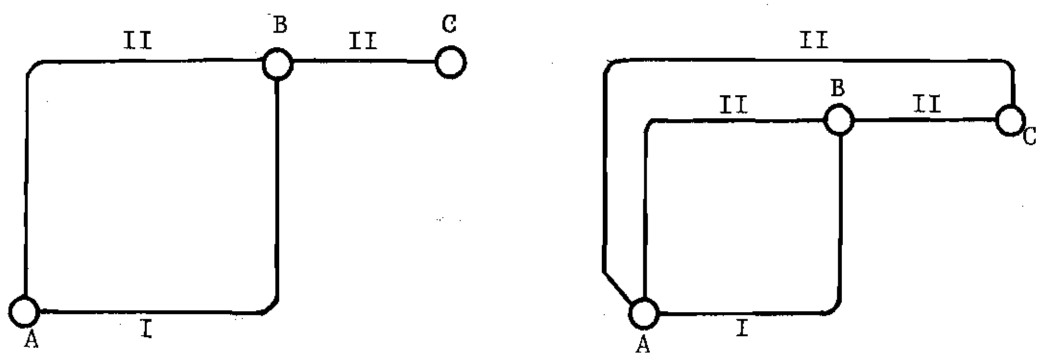

Figure 10. Construction of Revised Routes

From the diagram it can be seen that the shortest path from A to C passes through B in the original network but in the revised network the same path ignores node B since it is an intermediate stop. This revised network then, is the heart of the Le Clercq algorithm.

The shortest path finding procedure used here uses the same expanded network, but rather than consider it implicitly, the network is completely constructed in an explicit fashion. The network is expanded each time a shortest path tree is built, thus the arc travel times on the new network can be updated as a result of the variable travel times in the original network. A labeling procedure, specifically the Moore tree building algorithm is then applied to the expanded network and the shortest paths found. It is a simple matter to convert these paths to shortest paths defined in terms of the original network.

The algorithm used by Le Clercq could probably be converted to handle the variable arc times and find the paths more efficiently than the procedure used, especially for larger problems. A tree building algorithm already coded and ready for use on the Cyber was immediately available, however, so adding the procedure to expand the network was the only further work required to implement the shortest path algorithm used. This small amount of programming was a major factor in choosing not to convert Le Clercq's procedure.

One other consideration besides execution time dictates the need to use Le Clercq's procedure for large problems. Considering the expanded network explicitly means the entire network must be stored in the computer. The amount of storage required becomes quite large as many more arcs exist in the new network, than in the original. The small five node route problem below shows how the number of arcs increases at a rapid rate.

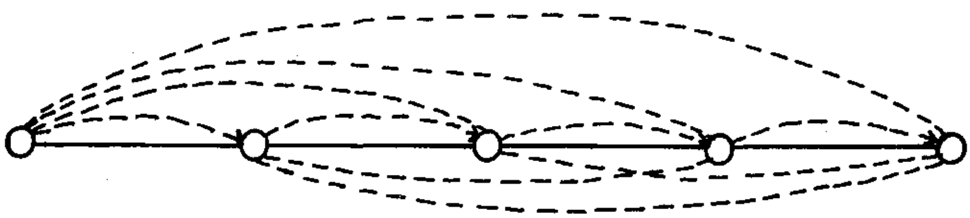

Figure 11. Revised Route Structure

The original network consists of four arcs and the expanded network includes 20 arcs, an increase of 2 1/2 times. In general, if the number of nodes in the route is n, and each node is connected as in the above diagram so there are n-1 two way arcs in the route, the expanded network would consist of N one way arcs where:

$$N = 2* \sum_{j=o}^{n-1} j$$

Thus the amount of storage needed grows exponentially with problem size, especially if long routes are being considered.

APPENDIX B

DEMAND AND TRAVEL TIME CURVES

This appendix discusses 1) how the commodity demand curve axis can be rotated and the scales changed to allow easy comparison with the path travel time curve, and 2) how calculating the equilibrium point of the demand-travel time curves is affected by paths containing several arcs for one commodity and by several paths of commodities sharing arcs.

Consider the demand curve and perceived travel time curve as developed in Chapter III. They are pictured below for convenient reference.

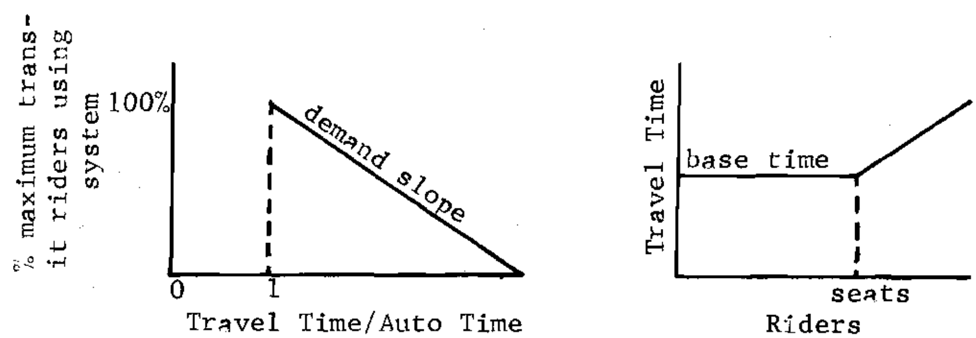

Figure 12. Demand and Travel Time Curves

The equilibrium point for the single commodity single path can be determined graphically be representing both curves on the same set of axis and identifying the point where the curves cross.

This is accomplished by first changing the scales on the axes of the demand curve. The abscissa is multiplied by the constant term corresponding to the auto travel time of the commodity and the ordinate is multiplied by the constant term corresponding to the commodity's maximum number of potential riders. This scale change alters the slope of the demand; the new slope is given by:

$$\text{New Slope} = \text{Old Slope} * \frac{\text{Maximum \# Riders}}{\text{Auto Travel Time}}$$

The old slope is the quantity required for input data to the solution procedure. Notice that the slope is always negative. When the axes are rotated making travel time the ordinate and the number of riders the absicssa, the demand curve slope becomes the inverse of itself. The resulting curve and the demand curve and travel time curve together are presented below.

In Chapter IV a simple example with one commodity having one path consisting of a single arc was examined. For the case where the number of arcs in the path may be greater than one, the same procedure can be used to determine the equilibrium point. This is because only the first arc is used to determine the number of seats; the other arcs contribute only to the seated travel time for the path.

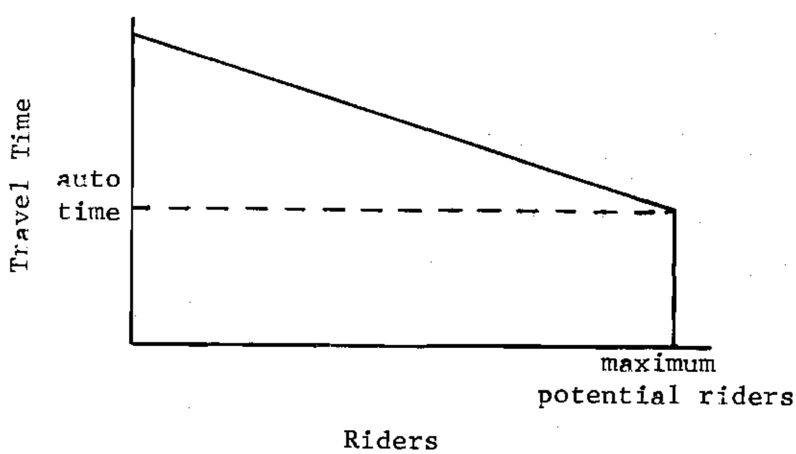

Figure 13a.

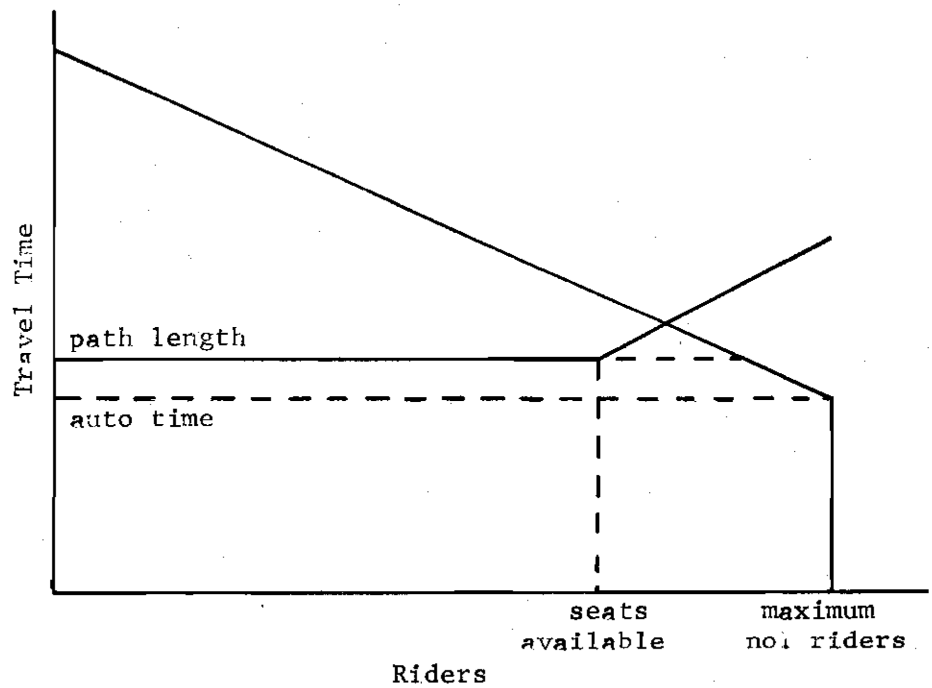

Figure 13b.

Figure 13. Expanded Demand Curves

This increase is reflected by the two travel time curves depicted below.

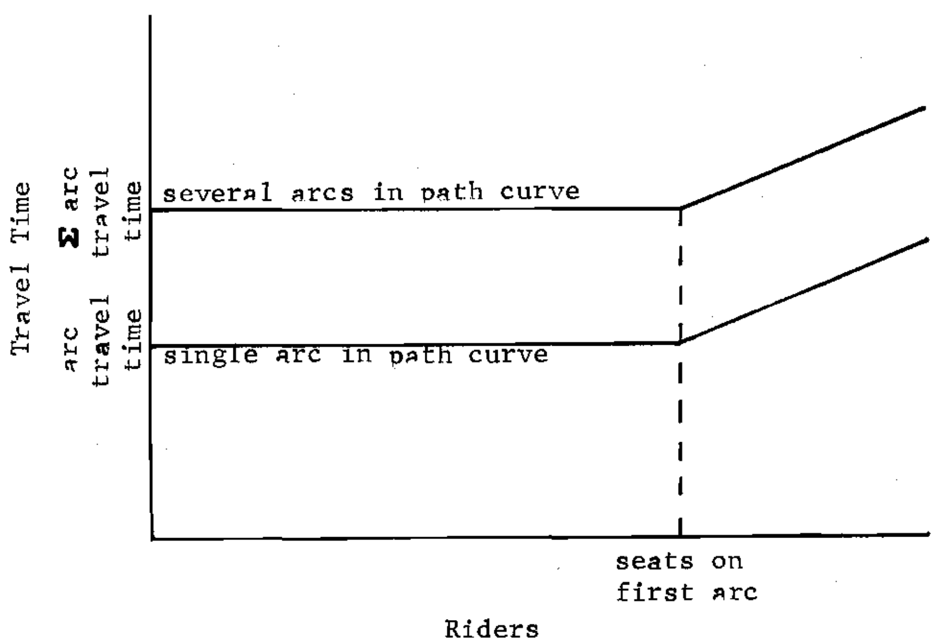

Figure 14. Travel Time Curves for Multi-Arc Path

The travel time curve will change in a more drastic manner when changing routes during a journey is considered. In these cases the initial arc in each route determines the number of seats available for part of the path. The travel time curve then is given by the equation:

$$\text{TRAVEL TIME} = \sum_{m=1}^{M} [\sum_{n=1}^{M_N} (\text{seated time } M_n + \max[0, \#\text{riders} - \#\text{seats}] M_1 * \text{travel time slope}]$$

where M is the total number of routes in the path and M_N is the number of arcs in the route M.

This is still a piecewise linear curve, with the changes in slope corresponding to the seating capacities of the various initial arcs on the routes taken, and the curve having a slope of zero until the number of riders exceeds the seating capacity of any route's initial arc.

Now consider the following example of two commodities in a network of three nodes and two one way arcs. Both commodities have the same origin node and only one possible path. The following information completely describes the example.

Commodity	Origin	Destination	Auto Travel Time	Maximum Riders	New Slope
1	1	2	5	10	-1
2	1	3	7	16	-1

Arc	No. Seats	Travel Time	
1	5	8	+1
2	5	12	+1

the following two graphs, one for each commodity, represent the path-demand curves.

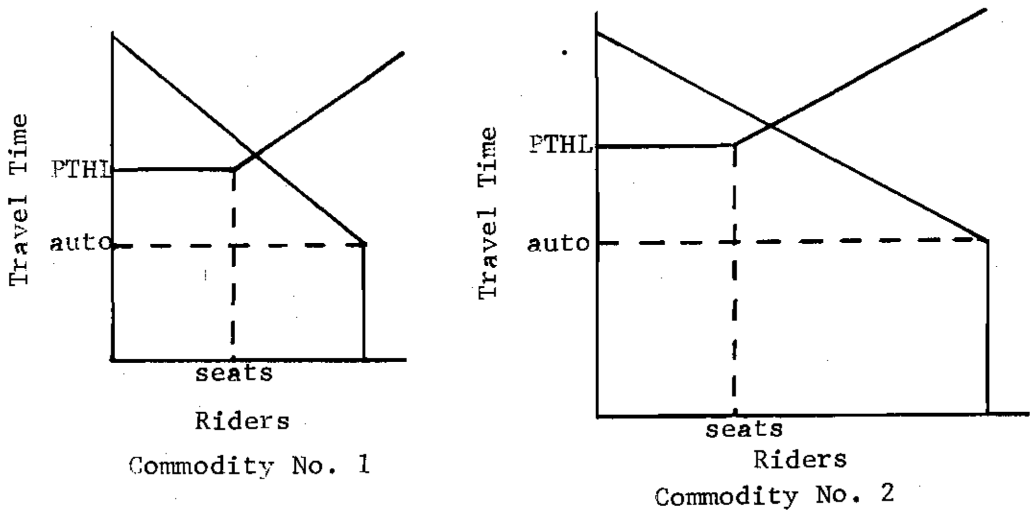

Figure 15. Two-Commodity-Partial Path Sharing Travel Time and Demand Curves

If each commodity is considered individually the number of riders respectively are 6 and 8, with travel times of 9 and 15. However, if all 14 riders are assigned to the paths, the travel times would change to 17 and 21. Therefore, both commodities must be considered simultaneously. Doing so yields an equilibrium of 3 and 6 riders with travel times of 12 and 16. Thus the simple graphical solutions no longer suffice, especially when there are large numbers of commodities taking paths consisting of more than one route and several arcs.

In Chapter IV the choice of an approximation procedure was discussed for use in the first two passes of the solution algorithm. In these two procedures, the emphasis is not on the calculation of demands but only estimates for use in finding shortest paths. The following graphical interpretation of the approximation procedure is presented to give further insight as to how the weighting scheme affects the final demand.

Consider a single commodity with a shortest path of several arcs along a single route which shares its initial arc with several other commodities originating from the same node. The travel time and demand curve for the commodity can be represented as in Figure 16.

In step one of the procedure, the path lengths before loading from this node are stored in array PTHLA. This point is marked in Figure 16. Then from this travel time, the corresponding point 1) on the demand curve is determined (DEMA) for all commodities. Then the travel time is again determined, this time with all of DEMA assigned (point 2 on Figure 16). Note how this demand is not the same as DEMA on the single commodity curve. The new path length is stored as PTHLB and the demand corresponding to this new travel time is stored in DEMB.

Figure 16 shows how the procedure moves from one point to another on the curves, and how the travel times and demands bound the final demand estimate. Also, it can be seen how all the commodities are considered, so the procedure is actually performing a modified spiral search.

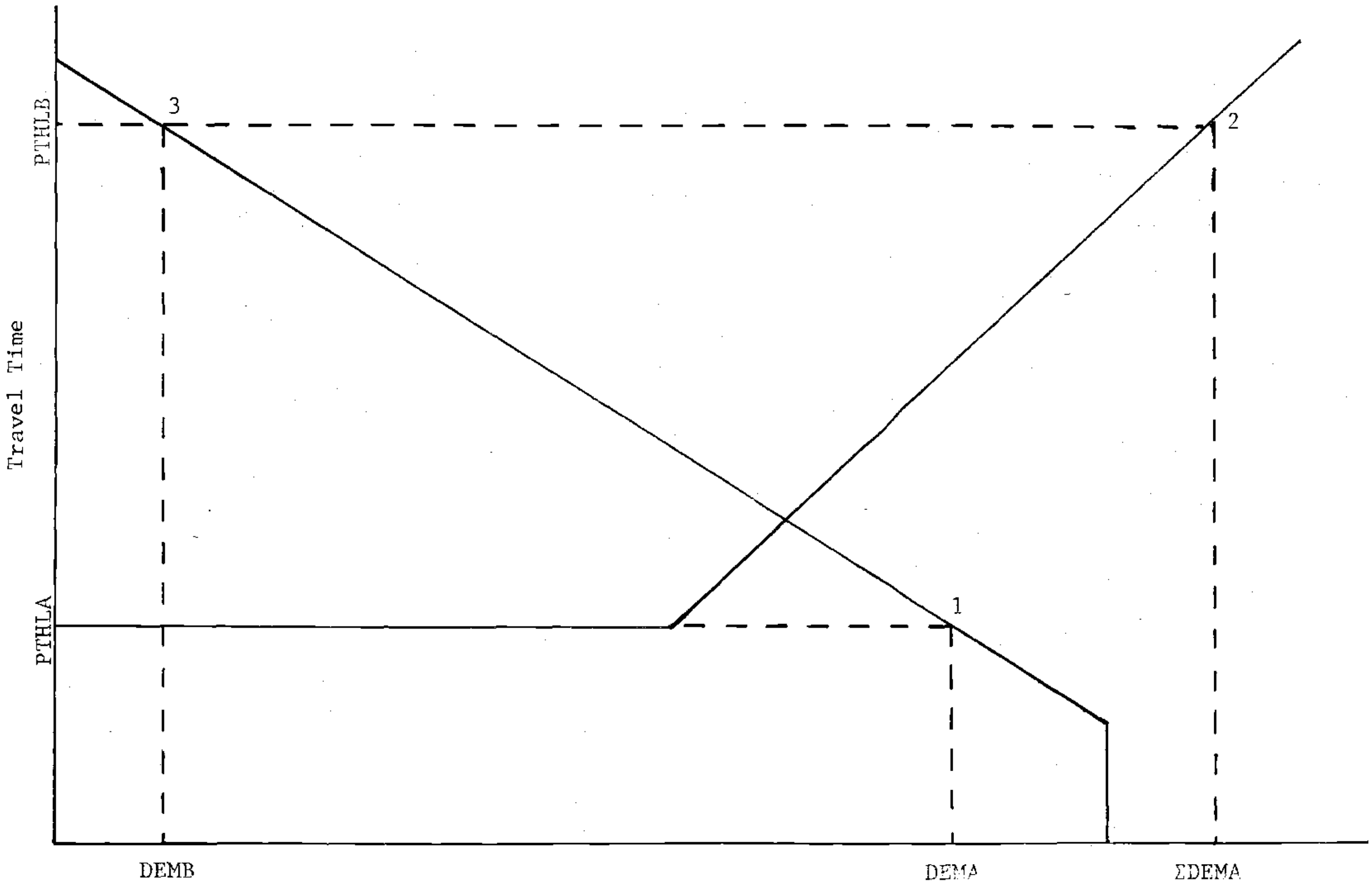

Figure 16. Demand Approximation Procedure

APPENDIX C

NETWORKS USED IN TEST PROBLEMS

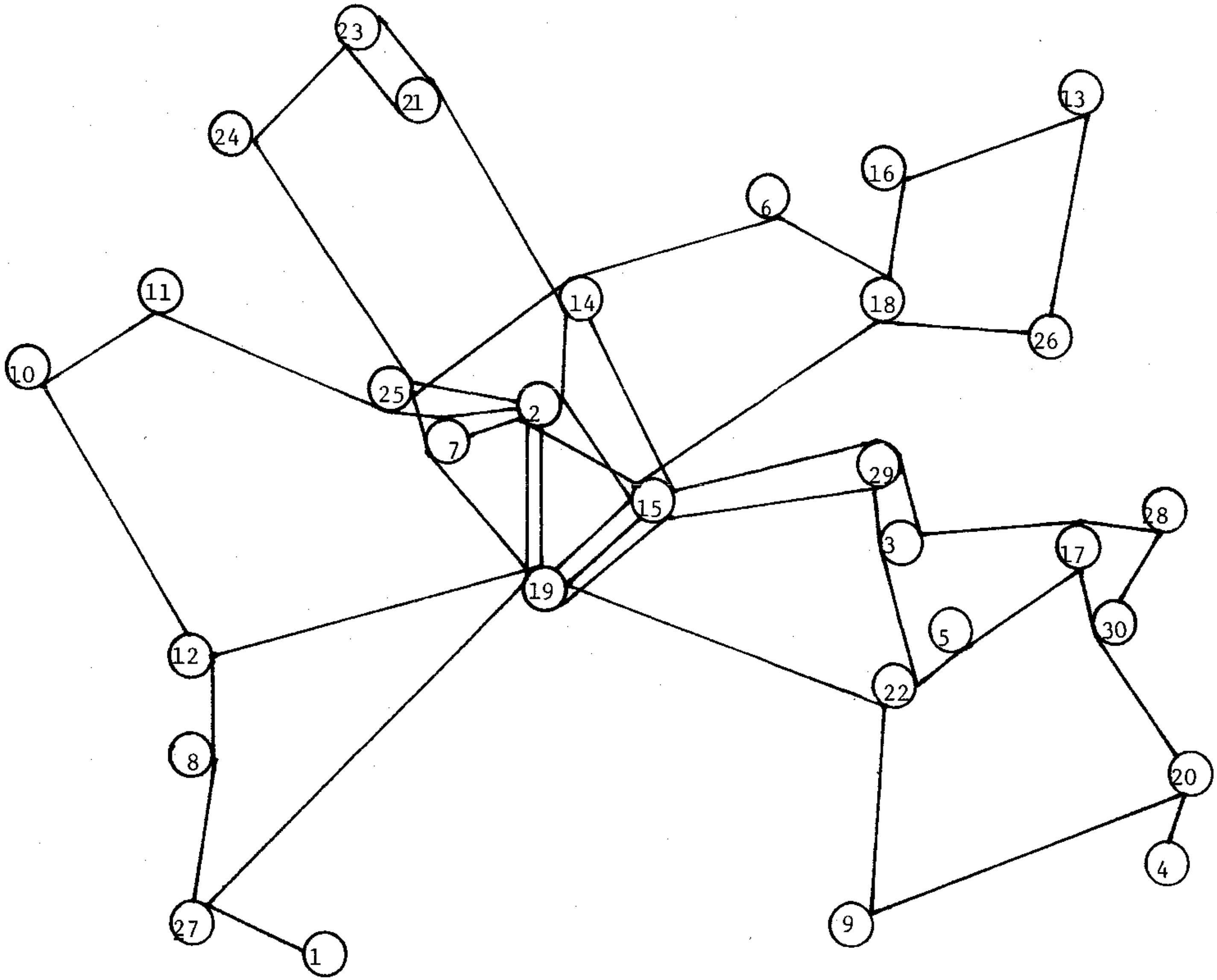

Figure 17. Network for Factorial Design Test Problem with 10 Routes

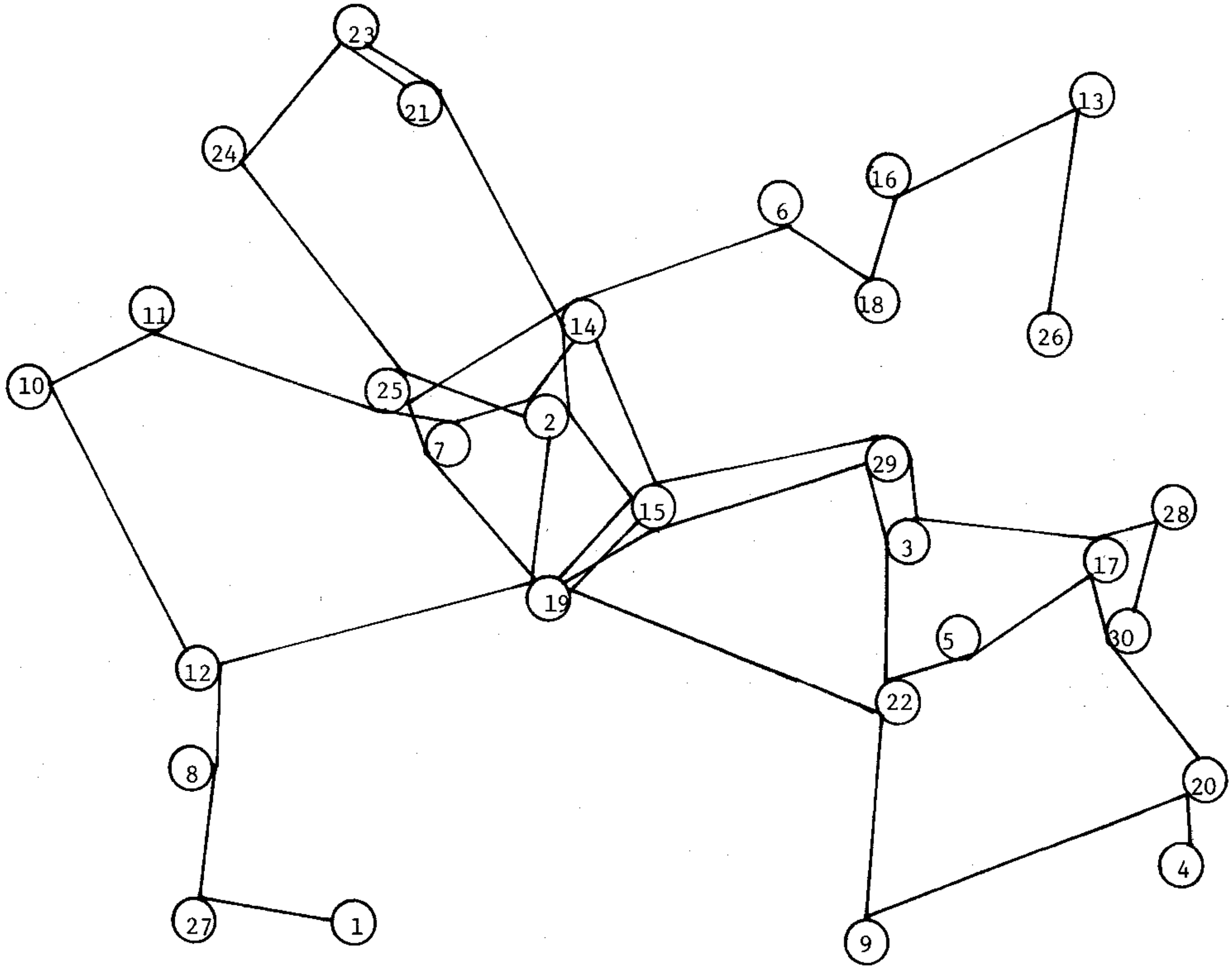

Figure 18. Network for Factorial Design Test Problem with 8 Routes

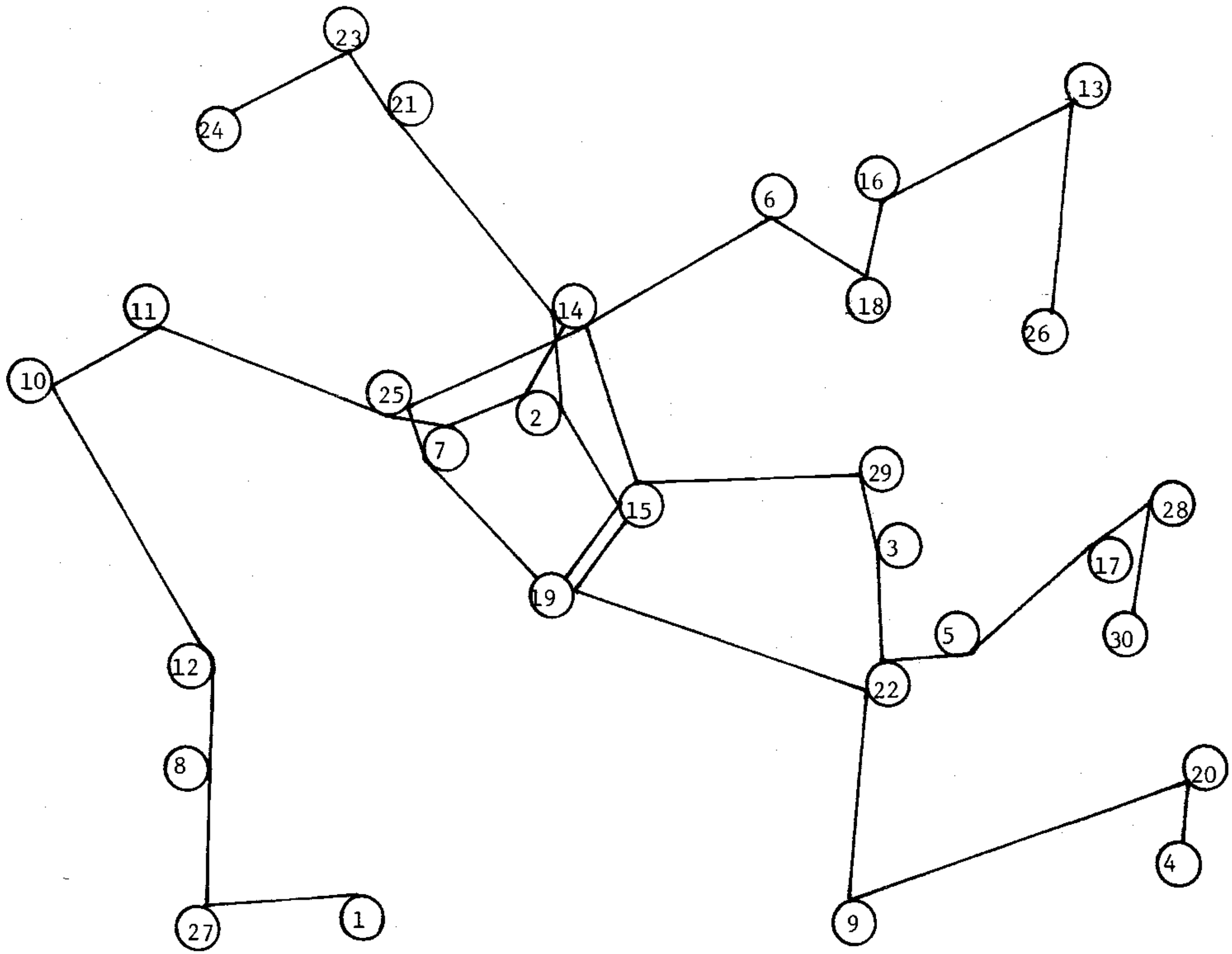

Figure 19 . Network for Factorial Design Test Problem with 5 Routes

Table 6. Potential Demand Matrix used as One Replication in Factorial Design

Origin Nodes	Destination Nodes 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	2	0	16	0	9	0	10	0	16	0	44	0	12	0	45	0	2	0	22	0	18	0	30	0	14	0	6
3	0	0	0	1	0	7	0	3	0	3	0	6	0	18	0	5	0	23	0	1	0	22	0	15	0	38	0	18	0	4
4	0	0	3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	3	0	0	0	1	0	0	0	2	0	0	0	0
5	0	31	0	1	0	0	6	0	4	0	6	0	8	0	76	0	5	0	38	0	0	17	0	10	0	23	0	11	0	3
6	0	0	7	0	0	0	2	0	2	0	3	0	2	0	12	0	1	0	5	0	0	3	0	6	0	10	0	4	0	0
7	0	12	0	0	5	0	0	2	0	1	0	2	0	6	0	2	0	6	0	0	3	0	4	0	6	0	2	0	15	0
8	0	0	3	0	0	2	0	0	2	0	7	0	2	0	19	0	0	7	0	0	3	0	4	0	9	0	1	0	7	0
9	0	0	5	0	0	3	0	1	0	0	0	3	0	7	0	2	0	8	0	0	3	0	3	0	6	0	0	3	0	0
10	0	10	0	0	2	0	2	0	0	0	4	0	2	0	6	0	0	8	0	0	5	0	5	0	10	0	1	0	7	0
11	0	0	10	0	0	4	0	4	0	3	0	0	3	0	20	0	1	0	12	0	0	10	0	12	0	12	0	6	0	1
12	0	0	6	0	0	2	0	3	0	3	0	0	2	0	37	0	2	0	16	0	0	5	0	6	0	7	0	3	0	0
13	0	10	0	0	8	0	2	0	1	0	4	0	0	11	0	2	0	13	0	0	5	0	4	0	8	0	0	3	0	0
14	0	39	0	2	0	7	0	3	0	4	0	7	0	0	30	0	3	0	14	0	0	7	0	6	0	10	0	11	0	2
15	0	0	44	0	29	0	12	0	7	0	23	0	7	0	0	10	0	36	0	1	0	14	0	13	0	65	0	33	0	6
16	0	0	5	0	0	2	0	1	0	1	0	1	0	11	0	0	2	0	9	0	0	5	0	3	0	8	0	0	11	0
17	0	0	3	0	0	2	0	0	2	0	0	2	0	6	0	1	0	0	4	0	0	3	0	1	0	4	0	0	7	0
18	0	0	23	0	15	0	7	0	4	0	9	0	5	0	34	0	3	0	0	2	0	20	0	15	0	32	0	18	0	4
19	0	0	15	0	12	0	5	0	4	0	11	0	4	0	26	0	2	0	0	0	6	0	7	0	37	0	5	0	39	0
20	0	3	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
21	0	21	0	0	6	0	4	0	2	0	5	0	3	0	17	0	2	0	7	0	0	0	9	0	18	0	1	0	21	0
22	0	0	9	0	0	3	0	1	0	1	0	2	0	7	0	5	0	24	0	0	5	0	0	4	0	12	0	6	0	1
23	0	0	7	0	0	3	0	3	0	3	0	4	0	7	0	2	0	19	0	0	10	0	0	9	0	9	0	4	0	0
24	0	19	0	0	5	0	4	0	1	0	6	0	2	0	11	0	1	0	16	0	0	6	0	0	16	0	2	0	15	0
25	0	0	14	0	10	0	6	0	3	0	11	0	3	0	29	0	2	0	15	0	0	7	0	7	0	0	5	0	39	0
26	0	31	0	1	0	4	0	3	0	3	0	8	0	31	0	9	0	31	0	0	12	0	9	0	20	0	0	10	0	1
27	0	0	2	0	0	2	0	1	0	0	3	0	0	4	0	0	0	4	0	0	2	0	2	0	6	0	0	2	0	0
28	0	14	0	0	5	0	2	0	3	0	5	0	3	0	31	0	2	0	12	0	0	7	0	3	0	11	0	0	0	0
29	0	43	0	2	0	6	0	3	0	2	0	5	0	17	0	6	0	21	0	1	0	7	0	12	0	33	0	16	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

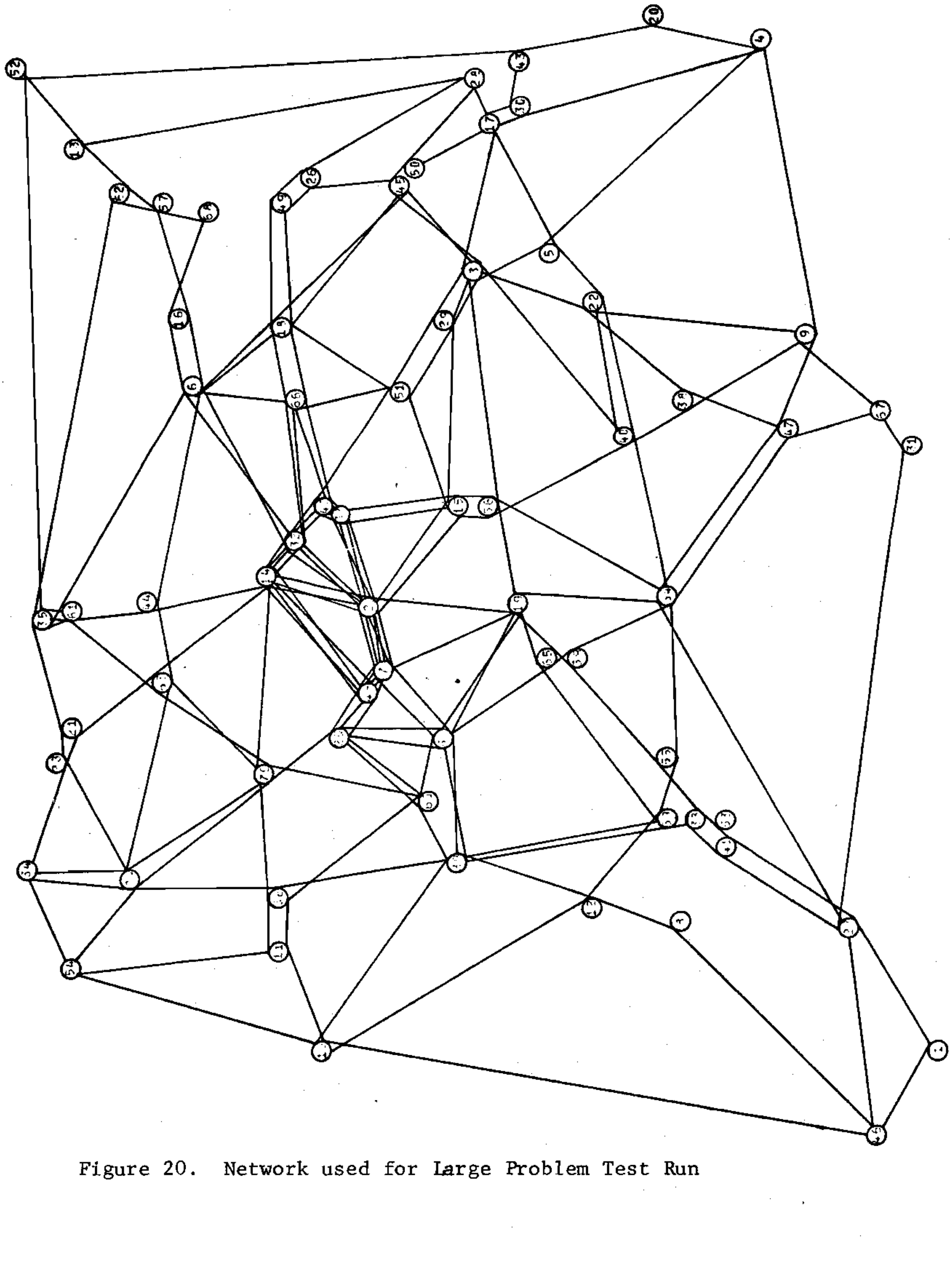

Figure 20. Network used for Large Problem Test Run

BIBLIOGRAPHY

1. Aburto-Avila, Jose Luis, Calson, Robert C., and Gray, Paul; "Estimating Demand in an Urban Mass Transportation System", Paper presented at the 41st National Meeting of ORSA, in New Orleans, La., April 26-28, 1972.

2. Chrique, Claude, and Pierre Robillard, "Common Bus Lines", Transportation Science, Vol. 9, No. 2 May 1975, pp. 115-121.

3. Dial, Robert B., "A Probabilistic Multipath Traffic Assignment Model which Obviates Path Enumeration", Trans. Res. Vol. 5, pp. 83-111.

4. David, Frank W., "The Role of Major Activity Center Mass Transit", Paper presented at XIX International TIMS Meeting, Houston, Texas, April 6, 1972.

5. Dobsen, Ricardo, Golob, Thomas F., and Gustafson, Richard L., "Multidimensional Scaling of Consumer References for a Public Transportation System: An Application of Two Approaches", Socio-Econ. Plan Sci. Vol. 8, pp. 23-26.

6. Fearnside, K. and Draper, D.P., "Public Transport Assignment-A New Approach", Traffic Engineering and Control, Nov. 1971, pp. 298-299.

7. Florian, Michael and Sang Nguyen, "A Method for Computing Network Equilibrium with Elastic Demands", Transportation Science, Vol. 8, No. 4, Nov. 1974, pp. 321, 332.

8. Florian, Michael, "On Modelling Congestion in Dial's Probabilistic Assignment Model", Transportation Res. Vol. 8, pp. 85-86.

9. Gensch, Dennis H., and Golob, Thomas F., "Testing the Consistency of Attribute Meaning in Emperical Concept Testing", School of Business Administration, The University of Wisconsin - Milwaukee.

10. Hoel, L., Evaluating Alternative Strategies for Central-Core Distribution", Highway Research Record No. 293, 1967.

11. Horton, Frank E. and Louviere, Jordan J., "Behavioral Analysis and Transportation Planning: Inputs to Transit Planning", Transportation 3 (1974) 165-182.

12. Hutchinson, B.G., "Principles of Urban Transport Systems Planning", Copyright 1976 by Scripter Book Company.

13. LeBlanc, Larry J., "An Algorithm for the Discrete Network Design Problem", Transportation Science, Aug., 1975, pp. 183-199.

14. Le Clercq, F., "A Public Transport Assignment Method", Traffic Engineering and Control, June 1972, pp. 91-96.

15. Martin, B.V., "Public Transport Assignment", letters to the Editor, Traffic Eng. and Control, Feb. 72, p. 455.

16. Meyers, J.R., Kan, J.F., and Wohl, M., The Urban Transportation Problem, Harvard University Press, Cambridge, Mass., 1965.

17. O'Faherty, C.A., and Managan, D.O., "Bus Passenger Waiting Times in Central Areas", Traffic Engineering and Control, Jan. 1970, 419-421.

18. Steenbrink, P.A., "Optimization of Transport Networks", Copyright 1974 by John Wiley & Sons Ltd.

19. Swoveland, Cary, "Decomposition Algorithms for the Multi-Commodity Distribution Problem, Thesis October 1971, UCLA.

20. Tomlin, J.A., "A Linear Programming Model for the Assignment of Traffic", Proc. 3rd Conference of the Australian Road Research Board, Vol. 3, Part 1, pp. 263-271 (1966).

21. Tomlin, J.A., "A Mathematical Programming Model for the Combined Distribution - Assignment of Traffic", Trans. Sci. 5, 120-150 (1971).

22. Tomlin, J.A., and S.G. Tomlin, "Traffic Distribution and Entropy", Nature 220, 974-976 (1968).

23. Watson, Peter L., "The Value of Time; Behavior Models of Modal Choice", D.C. Health and Co., Toronto, 1974.

24. Wardrop, J.G., "Some Theoritical Aspects of Road Traffic Research", Proc. Inst. Civ. Eng., 1, Part II, 325-378 (1952).

25. Yu, Jason C., and Hauthorne, Richard C., Assessing Urban Transit Systems: A Goal Programming Approach, Paper presented at the 54th Annual Meeting of the Transportation Research Board Washington, D.C., Jan. 1975.